An Analysis of Neural Networks and Time Series Techniques for Demand Forecasting

A thesis submitted in partial fulfilment of the
requirements for the degree of

MASTERS IN COMMERCE

at

RHODES UNIVERSITY

by

DAVID WINN

January 2007

ABSTRACT

This research examines the plausibility of developing demand forecasting techniques which are consistently and accurately able to predict demand. Time Series Techniques and Artificial Neural Networks are both investigated.

Deodorant sales in South Africa are specifically studied in this thesis. Marketing techniques which are used to influence consumer buyer behaviour are considered, and these factors are integrated into the forecasting models wherever possible.

The results of this research suggest that Artificial Neural Networks can be developed which consistently outperform industry forecasting targets as well as Time Series forecasts, suggesting that producers could reduce costs by adopting this more effective method.

ACKNOWLEDGEMENTS

I would like to thank my supervisors for their assistance on this thesis. Professor Burton for introducing me to a fascinating field of study, and aiding me in its understanding. Doctor Elliott for his excellent guidance and supervision and finally, Professor Radloff, thank you for all the guidance, advice, and especially for all the assistance at such short notice.

Thank you Bryony, for believing in me and encouraging me to undertake this project, and for your support throughout. A special thank you for all your proofing, aiding me in making it readable.

DECLARATION

This masters thesis represents my own work, and acknowledgement is given in the references wherever information is derived from another source. No part of this masters thesis has been or is being concurrently submitted for another qualification at another university

SIGNED....................................... **DATE:** 24 JANUARY 2007

TABLE OF CONTENTS

INDEX OF TABLES

ILLUSTRATION INDEX

Don't never prophesy: If you prophesies right, ain't nobody going to remember and if prophesies wrong, ain't nobody going to let you forget" - Mark Twain

CHAPTER 1
INTRODUCTION

1.1. Introduction

Demand forecasting is the process of predicting the number of goods to be sold in a future time period. Accurate demand forecasting is a complicated process which, done poorly, can add excess unnecessary costs to businesses. This thesis will examine different methods of demand forecasting with respect to the deodorant category of fast moving consumer goods, focusing specifically on those sold at Pick 'n Pay Superstores in the Western Cape in South Africa. Firstly, the thesis will explain the marketing theory which is used by companies to influence demand for a specific product. Next, it will highlight two different techniques of predicting demand, namely Time Series Analysis and Artificial Neural Network forecasting. These techniques will then be used on data sets of deodorant sales provided by Johnson and Johnson, South Africa. Finally, these results will be analysed, showing the techniques used to predict demand, produce better results in terms of consistent accuracy than the current industry expectations. This introduction will provided insight into the necessity of the factors discussed above and will highlight their utility to the research as a whole.

1.2. Problem Definition

Firms require sufficient products to satisfy consumer demands at all times. Having exactly the right number of products on the shelf is difficult at best and impossible at worst, and firms attempt to minimise this error. Minimising this error is important as the error carries a cost. This cost is having too much stock or too little stock. Too much stock results in expenses such as storage fees, insurance, as well as running the risk of pilferage, the stocks perishing or the product becoming obsolete (Morrell, 1967; Krautter, 1999). There is an additional cost as capital is tied up in inventory which could be used more profitably in other activities (Berkowitz et al., 2000). Stock shortages also carry a cost, as consumers cannot purchase the product, thus the firm immediately loses revenue. This also carries additional risks as the consumer may switch to another brand for future purchases (Corstjens and Corstjens, 1999), thus the firm losing future revenue. Firms attempt to diminish these costs by predicting consumer demand as close as possible to the actual demand.

1.3. Aims of the Study

This research aims to provide an understanding of the efficacy of two models of demand planning and forecasting. The models are Time Series Analysis and Artificial Neural Networks. The research is conducted within the South African Deodorant Category in the field of demand forecasting. It identifies the factors which marketeers have at their disposal to influence the consumer decision process to encourage the use of certain brands and uses this information for the construction of more accurate models. The research shows that a greater understanding of marketing factors that influence demand are essential for more accurate forecasting. A Time Series Forecasting Model and an Artificial Neural Network Model Forecaster are developed and tested to determine whether demand forecasts can be predicted consistently and accurately. Ultimately, the development of an accurate demand forecasting model will enable firms to reduce costs through a reduction in inventory management expenses. The research conducted for this thesis shows that the development of such a model may indeed be possible.

1.4. Chapter Content

This thesis consists of eight chapters, which focus on demand planning, marketing theory, Time Series analysis, Neural Networks and finally results of the research. This section is designed to briefly explain the necessity and impetus for the foci of the thesis. It will explain the reasoning behind each chapter, and its importance to the research as a whole.

Demand Planning is the process of estimating sales in a future time period. There are various patterns which demand can follow. There are a number of methods of predicting this demand including regression analysis, moving averages, Time Series and Neural Networks. Understanding the factors which influence demand aids in predicting future demand.

Marketing is the process of attempting to change the demand patterns of the consumer. According to economic theory, there are six factors which bring about changes in demand: the price of related goods (substitutes), expected future prices, income, expected future income, population and consumer preferences (Parkin et al., 2000). From a marketing perspective, pricing can be controlled and consumer preferences can be influenced, thus they are deemed relevant to the development of a demand forecasting model.

Marketing theory recommends the marketing mix as a method influencing consumer buying behaviour. The generic theory of Kotler (1972) recommends four elements that can be manipulated for successful sales. These elements are product, price, place and promotion. This thesis will acknowledge the influence of marketing techniques by taking them into account when predicting demand using the Artificial Neural Network, and will show it to be superior to Univariate Time Series analysis because it is able to take into account Kotler's four elements as different variables. Regression, Dynamic Modelling, and controlled Multivariate Time Series, are not considered in this thesis.

Understanding the ways in which a consumer market may be influenced to adjust purchasing behaviour of consumers, results in an increased knowledge of what factors may be useful in determining consumer demand. This in turn can be used to predict the demand. Historical data provide a basis for an estimate of future demand, with the other factors suggesting further changes in demand. Using this information demand forecasting models can be developed to predict demand, and hopefully reduce inventory costs. The greater the understanding of the marketing factors, the greater the chance of accuracy in prediction. This thesis aims at constructing accurate predictions, and so emphasises the necessity of comprehensive marketing knowledge and understanding in the construction of models of prediction.

The thesis will focus on two models of prediction: Univariate Time Series models and Artificial Neural Networks.

Time Series models assume that future observations are a function of equally spaced previous observations. Time Series data are often inappropriate in its natural form as the trends it exhibits are often non linear making it difficult to forecast accurately. As such it is transformed to create a stationary model, from which forecasts can be generated. In instances with only one variable, univariate models can be developed and parameters estimated to create a model which can be used in forecasting. For models which cannot be transformed to stationarity, smoothing methods are utilised to generate smooth functions from which forecasts can be generated. Time Series analysis provides a model based on scientific understanding.

Artificial Neural Networks is a highly flexible method of modelling. This flexibility is derived from the ability to concentrate on multiple variables within a single Neural Network. Neural Networks can model an extremely wide class of functional relationships, particularly non-linear ones (Darbellay and Slama, 2000). A Neural Network takes continuous-valued input variables and 'learns' their relationship to continuous-valued target values, a method known as supervised learning. The Neural Network is data driven in that it learns only from the training data presented to it and has no underlying parametric model. This means that the model produced is only as good as the data used, so the choice of explanatory variables is as important as for any other approach (Church and Curram, 1996). However, flexibility has a price. There is no established method for identifying the Neural Network structure that can best approximate the function mapping the inputs to the outputs. As a result, tedious and time-consuming trial-and-error procedures are unavoidable (Darbellay and Slama, 2000). This thesis demonstrates a Neural Network developed by the author as an example of an effective way of predicting demand within an environment where multiple variables are considered.

1.5. Chapter Organisation

This chapter is followed by a literature review in Chapter 2 through to Chapter 5 which puts the "research into context by showing how it fits into a particular field. It is useful in identifying knowledge gaps, and developing a research problem, for identifying a theoretical framework, identifying issues and variables related to the research topic" (Terre Blanche and Durrheim, 1999). Chapter two discusses reasons for forecasting and the types of demand patterns which can be used to describe a demand process. This chapter suggests different forecasting methods and briefly describes regression and moving average forecasts. A summary of these methods is presented at the end of the chapter.

Chapter three details marketing theory with specific reference to the marketing mix and the 4P's of price, place, promotion and product as described by Kotler (1972). The chapter is organised by defining marketing and the marketing mix. This is followed with a section on price, explaining how price influences the consumer decision process and how this impacts demand and the total units sold. This is followed with a section on place where inventory management is reviewed and the importance of accurate demand forecasting. Promotions are the third section outlining how various methods of promotion influence consumers which impacts their demand. The fourth section defines a product, and how management of a product can result in changes in demand. The chapter is concluded with a summary of the marketing factors that can influence demand.

Chapter four describes Univariate Time Series Analysis. The first section reviews the characteristics of a Time Series through stochastic processes. The second section highlights methods of transformation which are necessary to make stationary data which can be used to make accurate forecasts. Different forecasting techniques are describe in univariate techniques, with smoothing methods and Box Jenkins methods being described as a means of creating forecasts. The chapter is concluded with a description of how forecasts can be evaluated for their accuracy and with a summary of Time Series techniques which can be used in forecasting.

Chapter five discusses Artificial Neural Networks. The design and structure of a Neural Network is described which develops into a description of the mechanics of an Artificial Neural Network. This is followed by a section on the Neural Network architecture which details the requirements for its development and the Back Propagation Technique which is used to create a Neural Network which has forecasting abilities. A summary of Neural Networks concludes chapter five.

Chapter six lays out the methodology utilised in this research. The population and sample used in this study is discussed. The methods used in creating a Time Series and Neural Network demand forecasting model are discussed resulting in a discussion of the evaluation methods in determining how successful the forecasts are. The chapter ends with a summary of the research methodologies used.

Chapter seven is where the results are presented. The main findings of the Univariate Time Series research and the Neural Network development are presented. The results are then compared against 39 data sets to determine their accuracy and consistency to determine whether the forecasts are reliable. The results are also compared against industry requirements to determine whether the forecasts meet levels of acceptability as required by industry. A section on factor analysis is presented which gives an indication of what factors are most influential in developing a demand forecasting model. This is followed with sections on limitations of the research as well as future research that can be conducted in this field. The chapter ends with a conclusion of the results.

Chapter eight presents a conclusion of this research.

1.6. Conclusion

This chapter has provided an introduction to the research to be undertaken in this thesis. It has outlined the need for the research, and has given an indication of how it is to be carried out. The factors which influence consumer expenditure patterns are explored and how this affects demand. Demand planning methods in Univariate Time Series techniques and Artificial Neural Networks are also presented.

CHAPTER 2

DEMAND PLANNING

2.1. Introduction

Demand Planing is the process of predicting the number goods to be sold in a future time period. It is done with a view to controlling inventories with reasons described in Chapter 3. This thesis aims to detail the development of a model for prediction of demand and to do this with greater accuracy of forecasting than is usual in industry. It is imperative, therefore, to have an understanding of common demand trends and the effect of demand. Forecasting can be separated into short, medium and long term as described in table 2.1 (Lewis, 1975; Al-Saba and El-Amin, 1999).

Forecasting Method	**Range of time period**	**Examples**
Short-term	Days - Months	Demand forecasting at retailer (days), wholesaler (weeks), manufacturer (month)
Medium Term	Months - Years	Economic Forecasting
Long Term	Years - Decades	Technological forecasting

Table 2.1: Type of forecast and related time period

2.2. Forecasting

A forecasting system treats the total demand in each forecasting time unit as a single item of data. As the forecasting time unit is extended (i.e. from a day to a week), the average value of demand per unit time increases, but decreases relatively the variability of individual's values around that average value, and hence improves the accuracy of forecasts. However, longer forecasting time units reduce the speed of response of a forecasting system to changes in the demand pattern. Choosing an appropriate time unit is a compromise between the two (Lewis, 1975). For example, predicting on a daily basis, an average of 10 unit sales. An error of 1 unit is a 10% error. If this is extended to 30 days, to have the same error, it will be incorrectly estimated by 30 units, which is less likely. However, if production occurs on a weekly basis, monthly forecasting does not offer as much useful information as a weekly or daily estimate may offer (Thomas, 1968).

It has been argued that the aim of short term forecasting is not to predict demand in advance exactly, but to minimise forecasting errors over a period of time. A forecast with errors scattered on either side of the actual demand is what is required. However, should over or under predictions prove more costly, the prediction could be artificially biased to the contrary. Residual error is the error between actual demand and forecasts that have been made on a basis related to such values (Lewis, 1975).

In a general sense, demand forecasting assumes that individual demand values are drawn from a statistical population whose parameters can be estimated, and is not unduly constrained by external influences. In practice, this is not possible, even if the constraints are that supply can never be negative or that production can never be infinite at the other extreme. As such all demand populations are composed of a basic stationary component, which is shown through a non varying level of demand. There may be a basic level of variation as shown in figure 2.1 (Lewis, 1975).

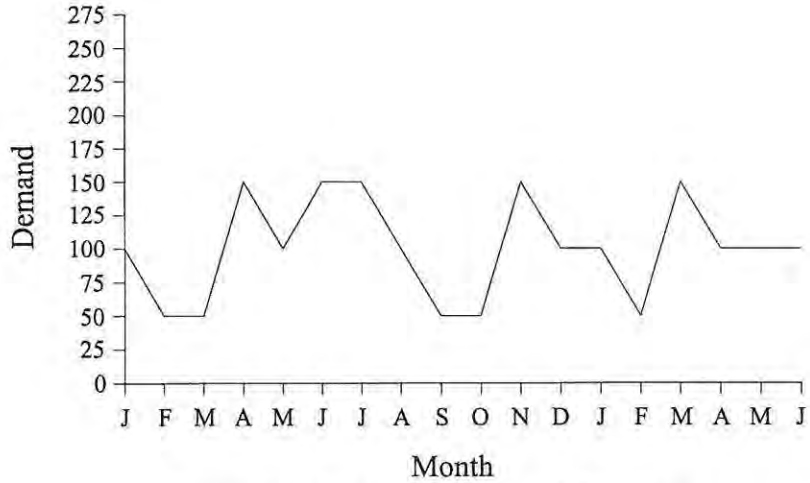


Figure 2.1: Stationary Demand Situation

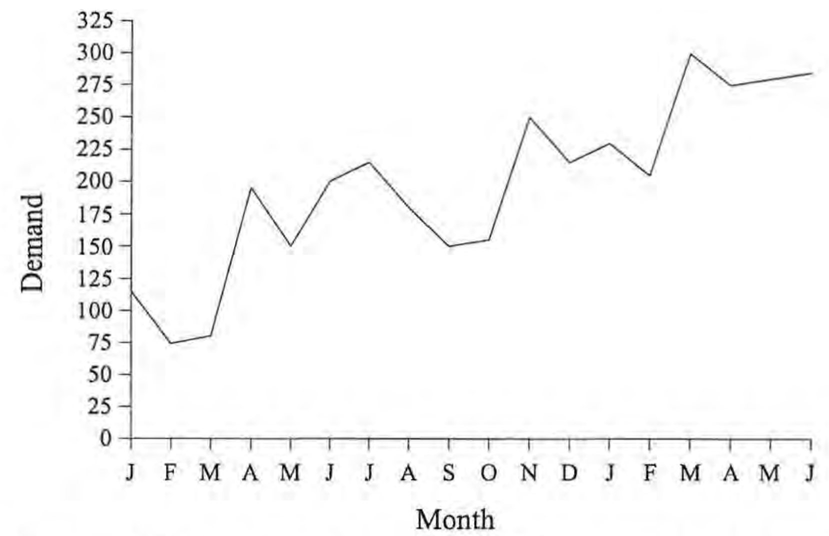


Figure 2.2: Stationary Demand with superimposed fixed growth

One may find a linear growth element superimposed on the stationary demand. In figure 2.2 there is a constant growth factor of 10 units, but with random variation in the demand. However, a fixed growth factor is unlikely, and a stationary demand with a random growth factor as shown in figure 2.3 is far more likely (Lewis, 1975).

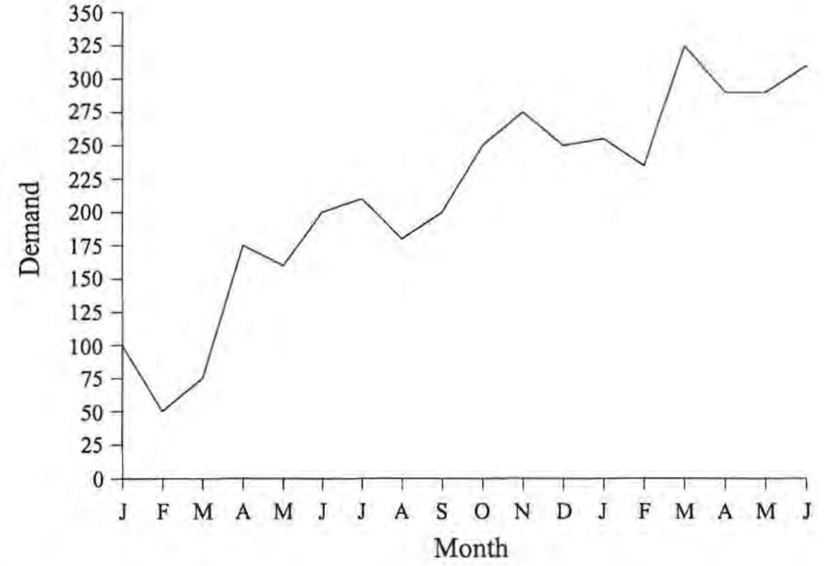


Figure 2.3: Stationary Demand with superimposed random growth

There may even be a seasonal element superimposed upon the stationary demand with random growth, which would be as shown in figure 2.4. Randomising the variation in the seasonal factor would complicate it even more (Lewis, 1975).

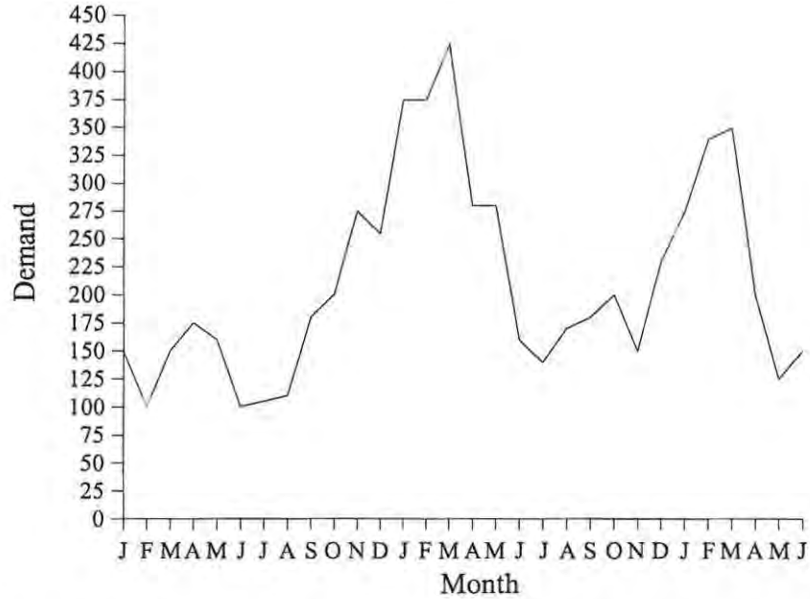


Figure 2.4: Stationary Demand with superimposed random growth and influenced by a fixed seasonal demand

The above graphs give an indication of how complicated a demand situation may be. However, the past data are known, so the object of forecasting is to attempt to establish what the underlying models may be, and decomposing those into models with parameters which may be estimated (Lewis, 1975).

2.2.1. Forecasting Methods

It is possible to study demands by plotting them and drawing lines, but this is time consuming, is dependant on the person who is physically doing it, and does not give any reliable measure of the accuracy. Proper statistical methods exist which can provide a free from operator bias, and can provide a consistent indication of the error of forecasts (Thomas, 1968). These include

- Regression analysis
- Moving Averages

- Time Series Analysis
- Neural Networks

Regression Analysis includes a graphical plot of past demands, and inserting a linear line of best fit to the data. Although this is relatively easy and may capture a long term trend, it is not sensitive to any random fluctuations or seasonal indices. As a result there is often a large error, which is caused by the random fluctuations, and the error in estimating the underlying model. As a result regression analysis is not regarded favourably for demand forecasting (Thomas, 1968).

Moving Averages are very simple to apply, and a common technique. It forecasts on the premise that the following period will be the same as the average of the previous n periods, where n is determined by the operator. The advantage is that a small window period can be used to reflect short term demand changes, and a larger window to reflect long term averages. This window size is based on the length of the forecast, and the level of fluctuation between prior demands. For example, if there is little random fluctuation, and a one month forecast is needed, the average of the previous three months is appropriate. The major flaws with this method are that it lags badly on any trends, and an isolated spike in demand will cause the average to be changed over the whole period it is in the moving average window. The moving average can also be weighted such that certain information is considered more pertinent, for example, greater importance can be placed on the demand from the previous month, as opposed to 10 months prior (Thomas, 1968; Lewis, 1975).

These two techniques are not explored further in this thesis, owing to the fact that they are generally accepted as inadequate forecasting procedures and more advanced techniques are now available. Time Series Analysis and the use of artificial Neural Networks provide the basis for the models explored in this research.

Time Series Analysis utilising exponential smoothing, and Box Jenkins smoothing is considered more reliable, and is explained in greater detail in Chapter 4. Neural Networks is a considerably newer technique, and has been shown to be successful in demand forecasting, and is explained in Chapter 5.

2.2.2. Demand Factors

Demand planning requires the ability to predict how markets will change with the change of a variable. Market response to price changes is affected by many elements including price sensitivity, competitive intensity, and product characteristics. In addition, how additional factors influence this is also relevant: a price discount with feature advertising fundamentally differs from one without advertising (Mulhern and Leone, 1995).

Understanding the interaction of the variables that influence demand is important, as this enables improved prediction. The influence of these factors will be further explored in the following chapter.

2.3. Conclusion

This chapter has examined the concept of forecasting and different models used to predict demand of sales. Four such techniques have been introduced; regression analysis, moving averages, Univariate Time Series Techniques and artificial Neural Networks. This thesis will use the concept of demand explained here to conduct research on Time Series Analysis and Artificial Neural Networks as methods of predicting sales of products in the fast moving consumer goods industry.

CHAPTER 3

MARKETING

3.1. Introduction

This chapter reviews the literature pertaining to the marketing theory relevant to demand planning. It will focus on the elements used to influence the consumer purchasing process. Kotler (1972) identifies product, price, place and promotion as factors integral to successful marketing and these are examined to understand how they affect the consumer buying decision, and how this knowledge can be used to maximise sales, as well as to make accurate demand forecasts. The chapter will examine the meaning of marketing, the importance of the marketing mix and finally the impact of market maturity on demand.

According to the American Marketing Association, marketing is "the process of planning and executing the concept, pricing, promotion and distribution of ideas, goods and services to create exchanges that satisfy individual and organisational objectives" (Berkowitz et al., 2000:9). Peter Drucker (in Simkin, 2000) furthers this by stating that "the aim of marketing is to make selling superfluous. The aim is to know and to understand the customer so well that the product or service fits them and sells itself. It is significantly more than the commonly believed advertising and personal selling". Simkin (2000) identifies seven core themes that appear in most marketing definitions that can also be considered to be the objectives of marketing:

- The ability to satisfy customers through a product or service that meets a customer's needs.
- The exchange of a product or service in return for a payment or donation.
- The need to create a competitive advantage over competitors.
- The identification of favourable marketing opportunities that can be exploited to develop the market and expand the market's use of the product.
- To remain profitable to enable a viable future for the organisation.
- To use resources shrewdly as to lower costs and maximise a business's market position.
- Aim to increase market share in priority target markets

Although these seven aims are indeed generic goals for individual firms, firms require different marketing strategies to remain competitive as they look to develop different markets. It is their ability to define themselves in a market which allows a firm to create a competitive advantage through marketing. Kotler (1972) clearly defines the axioms of marketing and provides a typology of the different components of marketing. This subdivision has served as the basic premise for marketing theory. The next paragraph will discuss this segmentation as crucial to marketing management.

One of the first and most important aspects of marketing management is to segment the mass market into relatively homogeneous sub-segments and then select one or more of these groups as target markets based on factors such as consumer preferences or the demographics of the consumers (Datta, 1996; Simkin, 2000). This process helps the understanding of the structure of the market and also in the identification of the customer. This will also highlight the similarities or differences between the actual customer and expected target market, thus allowing for changes in marketing practice (Goldsmith, 1999). The process of understanding the needs of the consumers, and developing a product to satisfy them is a part of the strategic marketing process.

Strategic marketing is the process of ensuring that the marketing pursued meets required objectives. When this occurs, the marketing is regarded as successful. Successful marketing relies on the presence of two or more parties with unsatisfied needs, the desire and ability to satisfy the needs, a method of communication, and something to exchange (Berkowitz et al., 2000). This leads to the establishment of a market, with the presence of buyers and sellers to satisfy the needs as discussed above (Ferrell and Hartline, 2005). Within the market, there is then an attempt to influence consumers to purchase specific products, which is achieved through the manipulation of the marketing mix.

3.2. Marketing Mix

"The marketing mix concept is one of the core concepts of marketing theory" (Rafiq and Ahmed, 1995) as it describes the controllable aspects of marketing. This was introduced into mainstream marketing theory by Kotler in 1972 (Brown, 2002) and is the generally accepted theory of marketing. The marketing mix consists of four controllable factors commonly known as the 4P's (Kotler, 1994; Berkowitz et al., 2000) and is considered to be the tactical marketing tool kit manipulated by most marketing practitioners (Simkin, 2000). The 4P's of the marketing mix;

product, price, place, promotion are detailed later in this chapter. These are the factors that can be manipulated by an organisation to have an effect on the behaviour of the target market.

Although considered to be the heart of contemporary marketing management, the theory of the 4P's has come under much criticism from Marketing theorists who have argued that it is not sufficient for modern marketing (Rafiq and Ahmed, 1995; Goldsmith, 1999; Constantinides, 2002; Hakansson and Waluszewski, 2005). The 4P model is not sufficient for services marketing as it focuses on tangible goods only, and as such does not allow for service products. The 4P's has thus been expanded to include "packaging, people or process" (Crittenden, 2005). Other models have been proposed, with Bennett (1997) arguing that the 4P's is narrowly focused on internal variables and so he suggests the five V's (Value, Viability, Variety, Volume and Virtue).

In addition to these there are a number of uncontrollable external factors that complicate market growth that can be grouped as social factors, technological, economic, competitive, and regulatory factors (Weber and Dholakia, 1998; Berkowitz et al., 2000). These are environmental factors and although these factors cannot be manipulated by the company, they may have a impact on the effectiveness of the marketing mix and are briefly considered below.

Social forces include the demographic characteristics of the target population such as population size, as well as the values associated with ethnicity or culture. All of these factors may affect the way a product is viewed. Economic forces relate to the income, expenditure, and resources that affect the running of businesses and households (Ferrell and Hartline, 2005). Factors such as consumer income, disposable income[1], and interest rates[2] all affect the availability of finances for the firm to be used in developing products, and for consumers to purchase the products. Technology is particularly relevant especially in the modern environment. It refers to innovations that improve existing products, and other innovations that can replace existing products and companies (O'Malley et al., 1999; Ferrell and Hartline, 2005). Firms that produce products that satisfy a similar market can be seen as competitive forces. Regulatory forces are most commonly government restrictions, primarily to protect the consumers as well as companies from unfair and unethical practises. Most of the regulation put into place by Government is to ensure fair business practice, and safety issues for both the consumer and producer (Ferrell and Hartline, 2005). A

1 Disposable income is finances which are readily expendable by consumers

2 A decrease in interest rates allows more money to borrowed at a cheaper rate, which encourages expenditure

successful firm will continuously monitor the external environment, and be prepared to adjust to meet any new demands. Although these external forces cannot be controlled, the internal marketing mix is substantially easier to manipulate (Maxwell, 1998), and where the market competitive advantage is derived.

Despite the criticism examined above, and despite the other factors that can be considered important too, the 4P's has remained integral to marketing theory, as it remains a generally accepted basis on which marketing theory can be developed. Because of this, this thesis will focus on the 4P's in its discussion of marketing as a variable in relation to demand planning. The internal marketing mix, and the 4P's are detailed below.

3.2.1. Price

Price is the amount of money or currency required to facilitate an exchange and the only component of the marketing mix to generate revenue (Shipley and Jobber, 2001). It represents to the consumer the "economic outlay that must be sacrificed to engage in a given purchase transaction" (Lichtenstein el al., 1993:234). Literature generally is in agreement on the importance of price on the consumer buying process (Monroe, 1973), although it is noted that there are low levels of price comparison shopping which may suggest a lack of price sensitivity (Grewal and Marmorstein, 1994). A number of objectives as well as constraints need to be taken into account when the price of a product is set.

3.2.1.1. Pricing Objectives

The objectives of price setting could include profit maximisation, revenue maximisation, turnover maximisation, business survival or even corporate social responsibility (Lancioni, 1988; Berkowitz et al., 2000). Firms wishing to maximise revenue would balance the price against the number of units sold, as revenue is quantity sold multiplied by the price (Jolson and Hise, 1973). Ideally a firm would maximise both price and quantity to maximise total revenue, but that is rarely practical, as there is trade off between quantity demanded and price, which is further explained as the basic laws of supply and demand[3] (Ferrell and Hartline, 2005).

Survival pricing is where the price is set purely to generate the required cash flow to sustain the

3 An increase in the demand of a good results in a increase in quantity and a increase in price, while the increase in supply increases quantity, and decreases price (Parkin et al., 2000)

business (Berkowitz et al., 2000). Price may also be set with a social responsibility as the objective, where larger sales and profits are sacrificed for the benefit of the community. This form of pricing is the least common, and firms ordinarily wish to increase their profits. Profits may be maximised by attaining a desired return on investment, by maximising current profits, or obtaining a targeted turnover (Lancioni, 1988).

In an effort to maximise turnover and market share (ratio of product sales to total industry sales), firms may adjust the price to sell more units, thereby increasing turnover, and increasing the total share of the target market. When adjusting prices, firms should be aware of the elasticity of the product and the effect that price adjustments will have on demand. This is referred to as price elasticity and discussed below.

3.2.1.2. Price Elasticity

When adjusting prices in an effort to increase sales, firms should consider the elasticity of the brand or product. Elasticity refers to the responsiveness of demand to changes in product prices (Jolson and Hise, 1973) with the formula shown in 3.1.

$$\textit{Price Elasticity of Demand}\ (E) = \frac{\textit{Percentage Change of Demand}}{\textit{Percentage Change of Price}} \tag{3.1}$$

An elastic demand is when a small percentage change in price leads to a relatively larger percentage change in quantity demanded, and occurs when $E<1$. This commonly occurs where there are many competitors or substitutes of a particular product. Inelastic demand occurs when $E>1$, and a large percentage change in price leads to a relatively smaller percentage change in demand. This occurs when there are few competitors and the product is viewed as important, or non-substitutable. Staple foods and primary services fall into this category. Elasticity is important to consider when price setting occurs, as if the product is elastic ($E<1$), a small decrease in price may have a significantly larger affect on total revenue as a larger number of the product is sold (Berkowitz et al., 2000; Ferrell and Hartline, 2005). Krishnamurthi and Raj (1991) have found that the loyalty of customers also affects elasticity, with loyal customers being less price sensitive than customers who sample many brands. This also allows firms to adjust prices based on the loyalty of the customer base in the knowledge of how likely they are to switch to another brand. Balancing prices against elasticities and corporate objectives are not the only factors to consider when determining pricing. Other pricing determinants will now be discussed below.

3.2.1.3. Pricing Determinants

Firms suffer from pricing constraints which limit and control the prices charged. There are a number of constraints that limit the price that can be charged. These can be separated into regulatory constraints and natural constraints and are discussed below(Berkowitz et al., 2000).

Regulatory pricing is purely legislative where laws prevent artificial prices being put into place. Artificial prices can be instituted through price fixing, geographic, or discriminatory pricing where different regions or groups are charged differently (Ferrell and Hartline, 2005). In addition predatory pricing is illegal. This is where prices are made deliberately low with the intent of forcing another firm out of business (Berkowitz et al., 2000). Thus government tries to encourage fair market practises, where consumers are charged fairly, while ensuring that businesses cannot compete on price with the intention of forcing firms out of the market to create a monopolistic market. Although Government aids the fairness of pricing, most pricing is constrained by natural determinants.

Natural determinants are market related pricing factors. The basic laws of supply and demand heavily influence the price of goods or services, where increased prices result in a decrease in demand, and similarly, a decrease in price, results in an increase in demand (Ferrell and Hartline, 2005). As such firms attempt to balance prices against the expected demand at different pricing levels. The product class[4], or brand[5] also influences the price. Products with a strong brand can command higher prices as it has a known quality and reputation that customers are prepared to pay more for (Erdem et al., 2002).

The stage of a product in its life cycle (see 3.2.4.1) also affects the price (Lancioni, 2004). A new product can command a higher price when it is considered to offer additional benefits over older products. If the product becomes a "fad", the price may increase further. As the product becomes older, less innovative, and with increased competition, the price will decrease (Ferrell and Hartline, 2005). As the number of products in a product line (see 3.2.4) increases, the prices becomes more stable and representative, as there is a comparative basis against completing products.

4 A product class is a group of similar products which serve a similar function

5 A brand is defined as a name, term, sign, symbol or design, or a combination of them which is intended to identify the goods and services of one seller or group of sellers and to differentiate them from those of competitors (Kotler, 1998).

Firms should have a clear understanding of their pricing objectives and pricing alternatives, as there may be an additional cost in changing prices to adapt to new strategies. This cost of price changes is also a natural pricing constraint. Where products are sold on a tender process and only a few buyers are involved, the price can be changed easily and cheaply. However, products listed in numerous places incur a large cost when the price is changed as there is a cost in the actual process of changing the price, such as shelf prices, or in a catalogue where the price is more fixed (Berkowitz et al., 2000).

The firm's cost structure also affects the price as the firms needs to at the very least break even, such that revenues equal expenses. However, the cost structure should not be the main force behind pricing, as firms may run the risk of setting their prices too low. If a firms prices are based purely on their cost structure, yet their costs exceeds other firms, their prices will be set relatively higher which would result in customers migrating to the cheaper equivalent. Conversely, a product may be cheap to produce, but does not necessarily have to be sold at a low price (Ferrell and Hartline, 2005).

Competitors also act as a natural pricing constraint as the company has to consider both competitors products as well as substitute products[6] in the pricing decision. Prices should be set as to remain competitive against completing products and substitutes. The price is also constrained by the market in which the product competes. In purely monopolistic markets, price is set purely by the seller as there is no competition to aid in determining a fair price. In an oligopoly there are very few sellers, and these sellers try to avoid price competition, and as such the market often follows the market leader in determining the price. In monopolistic competitive markets there are many sellers who compete on non price factors. As such, price is set according to product differentiation where products have different characteristics which are emphasised which may be attractive to consumers (Berkowitz et al., 2000; Ferrell and Hartline, 2005). The final form of market is a purely competitive market where there are many producers and the price is determined by the market, so that producing firms are price takers (Hakansson and Waluszewski, 2005). These constraints to pricing affect the approaches taken in determining price. Some such approaches will now be discussed.

6 Products which are different yet fulfil similar functions, such as butter and margarine

3.2.1.4. Pricing Approaches

Having understood the environment and constraints under which a firm operates from a pricing perspective there are a number of different approaches that can be used to set prices. There are four different pricing approaches that can be used; demand, cost, profit, and customer orientation (Shipley and Jobber, 2001). Overlaps in the approaches occur, and these can be used concurrently for an effective pricing strategy to encourage further sales (Berkowitz et al., 2000).

A demand orientated approach places greater emphasis on buyer factors, such as consumer tastes and preferences and bases pricing according to the demands of the product. This is achieved through the product satisfying the needs of the customers. There are numerous pricing tactics such as 'skimming pricing' where a new product or innovation is introduced, and the price is set at the highest that the customers are willing to pay so that the firm maximises short term profits (Hart and Tzokas, 2000; Shipley and Jobber, 2001; Gupta and Benedetto, 2006). These customers are not price sensitive, as the product, and its quality are highly valued (Ferrell and Hartline, 2005). Firms may want market share as opposed to short term profits, and utilise 'Penetration pricing' where a low initial price is set on a new product to have immediately appeal to the market and obtain a large market share (Hart and Tzokas, 2000; Shipley and Jobber, 2001; Gupta and Benedetto, 2006). This aids in driving down per unit costs as sales volume increases due to economies of scale. Penetration pricing works particularly well in markets that are price sensitive as a lower price will result in significantly greater sales (Berkowitz et al., 2000). Contrary to penetration pricing, firms may use 'prestige pricing' where price is seen by the consumer as a representation of quality (Lichtenstein el al., 1993), a high price indicates high quality, and if the price falls, demand falls too, as the quality is deemed to have lowered as well. Simpler pricing techniques are also used, such as price lining or odd even pricing (Ferrell and Hartline, 2005), where prices are set at step levels in the belief that demand is inelastic between the levels. Odd even pricing is used to make the product appear cheaper, such as R99.95 as opposed to R100.00. The product appears cheaper and this may encourage the consumer to make a purchase. Firms which offer a range of products also engage in bundle pricing where two or more products are grouped in a single package (Ferrell and Hartline, 2005), offering a net price reduction to the customer, but generating increased sales for the producer. All of the above suggested pricing methods are dependent on their being a demand for the product, and producers manipulating the pricing structure to influence the demand and total units sold. These methods operate without a focus on the costs to produce the product.

Cost orientated pricing is where the firm focuses pricing decisions on the supply side of the business (Shipley and Jobber, 2001). There is an attempt to set prices according to a relationship with the total costs. The final price comprises the production costs of the product, marketing costs, overheads, and the profit margin. Specific cost orientated pricing methods include standard mark-up pricing, where the cost of a product is determined and a standard percentage increase is applied to obtain a profit. A more variable pricing technique is experience curve pricing where the price decreases as production experience increases and becomes more cost effective, thus a product may initially be highly priced, and gradually become cheaper over time (Berkowitz et al., 2000). Cost orientated pricing ignores competitors which may result in the firm's products being inappropriately placed in the market which is a disadvantage of this technique (Shipley and Jobber, 2001). Although focusing primarily on costs, cost orientated pricing still ensures that a profit is added to the pricing structure. Other methods may be used when firms have more rigorous profit requirements.

Firms wishing to ensure specific profit margins may engage in profit orientated pricing where a specific Rand value of profit or percentage of sales investment is required. In such cases, target pricing may be used, where a specific Rand value of profit is desired, so the final price is calculated based on the costs, volume expected to be sold, and the required net profit. Target return on sales or investment is where a specific percentage return is required, and the price is manipulated to achieve this (Berkowitz et al., 2000; Lamb et al., 2002). Again competitors are often ignored in this pricing method, which can be a major disadvantage, as the price may be set above comparable products, resulting in customers purchasing those.

Given that the three other pricing methods mentioned neglect a competitive comparison in pricing, a competitive orientation in pricing is beneficial too. Competition orientated approaches base the price on the market, which may include both competitors and consumers (Shipley and Jobber, 2001). Firms may utilise customary pricing where a traditional price is maintained, and other variables are change accordingly, such as the size of the product (Berkowitz et al., 2000). This ensures that a product remains a competitive price while other products may increase in value. Firms can use pricing to reflect a position in the market relative to their competitors, such as 'above, at, or below' the market price. Firms not only wish to state a position in the market, but also to attract customers and earn market share. To achieve this, firms may engage in loss leader pricing where a product may be sold at a loss to attract customers into a store, so that they other products are bought as well (Berkowitz et al., 2000).

Competitive pricing and demand pricing engages with customers on more of a psychological level than the other pricing techniques, as it tries to influence the consumer with their price relative to competitors and price offerings. The other techniques of cost and profit orientation set prices with the goal of obtaining profits rather than necessarily influencing consumer decisions.

3.2.1.5. Pricing Summary

Pricing is considered to be significantly important for the success of a product, especially in industries where there are few product characteristic differences, and is highly competitive. With price elasticities being typically larger than other elasticities in the marketing mix (Zenor et al., 1998), it is critical to get the right price mix, to maximise sales, while balancing this against the product's demand, as the price should be low enough to attract customers, but high enough to cover costs. Pricing is central to the marketing mix, as it is the most flexible component, and is the component of the 4P's that can be adapted most rapidly to react to competitors, or new market conditions (Shipley and Jobber, 2001). The immediate and significant impact of price changes on market shares and profits, as well as the potential for strong reactions from consumers and competitors, makes it a particularly important component for managers to be aware of in the marketing process (Srinivasan et al., 2000). Price does not create a competitive advantage, but rather provides a product a temporary advantage over products of similar features and quality, again emphasising the importance of appropriate pricing (Corstjens and Corstjens, 1999). As the entire marketing mix has an influence over products sold to consumers, place, promotion and product are discussed below to develop a full understanding of how each marketing mix element can influence the total units sold of a product. This understanding will further aid in the development of a demand forecasting model.

3.2.2. Place

Place involves a myriad of activities, but is mainly determining the most appropriate location, or outlets for the physical product to be sold or distributed. It also includes the problem of getting the product to this location, which is called supply chain logistics (Svensson, 2002). In addition to determining the correct location and logistic process, it is important to ensure that the sufficient stock levels are maintained to meet consumer demand (Berkowitz et al., 2000).

In determining the physical location for the product to be sold, the major consideration is that it is a place that the target market is likely to frequent (Lamb et al., 2002). Store locations can be broadly categorised as being in the Central Business District, regional shopping centres, a community shopping centre, strip locations, (cluster of stores), or in a power centre (huge shopping strip with multiple stores) (Berkowitz et al., 2000). Different consumers are likely to frequent the different stores, thus it is important that the product being sold has a presence in the different stores. The process of getting the product on the shelves in the different stores is commonly known as supply chain logistics.

3.2.2.1. Supply Chain

Supply chain logistics is more than the delivery of a product to stores, but rather a sequence of activities that firms perform that are required to create and deliver a good or service to consumers and industrial users. A supply chain encompasses the process from producers of raw materials, manufacturing, packaging, through to the final retailer. This process of getting the product on the shelf is vital, with customer disappointed and promotion expenditure wasted if the desired product is not available to be bought (Berkowitz et al., 2000; Lamb et al., 2002). The supply chain is not only important because of the production process, but has an impact on the marketing mix as well, and as a result can be managed to aid in increasing total sales (Svensson, 2002).

Goods are often classified according to the duration of the life cycle. Perishable goods, or ones with an urgent need for delivery, such as newspapers, require an extremely efficient supply chain as expired goods cannot be sold. In addition, the physical shape and size of the product is important, large bulky packaging enables fewer products to be fitted in transport vehicles and warehouses which increases transport and storage. Pricing factors are considered when products are sold in bulk, as this shortens the supply chain. Promotions also have an interaction with logistics as they must be coordinated with logistic activities to ensure product availability at the appropriate time. In addition logistics need to get the product to the right place at the right time in a condition which is acceptable to the consumer (Berkowitz et al., 2000). In ensuring this, numerous activities are required in the supply chain process to ensure its effectiveness. These include transportation, warehousing, packaging, materials handling inventory control, order processing, customer service levels, production, warehousing and returned goods handing (Berkowitz et al., 2000). However, the successful implementation of this is expensive. Costs can be reduced with more efficient inventory management.

3.2.2.2. Inventory Management

Successful inventory management requires a balance between stock shortages or excess stocks. A shortage in inventory may result in poor service with stock outs leading to customers switching brands, resulting in a loss of market share. An oversupply of inventory leads to higher costs as capital is tied up in inventory, in storage, and products run the risk of obsolescence (Krautter, 1999; Berkowitz et al., 2000).

Traditionally large inventories have been justified as a buffer against supply and demand fluctuations not predicted for in forecasting. Maintaining these inventories allows better service to be provided for customers in satisfying their demand for the products. Production in large volumes results in lower per unit production costs which is an additional justification (Hulbert et al., 2003). Large inventories also act as a hedge against supplier price increases as well as acting as protection for the firm against strikes and shortages as the firm will be able to continue to supply the product despite the problems. Although large inventories are beneficial, they are expensive to maintain and these costs are often very difficult to isolate. A number of these costs are described below. A large inventory leads to inefficient use of capital, as capital is tied up in inventory, where it could be used for other more profitable investments. Inventory is insured, and requires storage, with both being costly as well (Morrell, 1967). Finally, keeping inventory in storage results in other possible costs, such as damage, pilferage, perishability or obsolescence (Morrell, 1967; Berkowitz et al., 2000) . These inventory related costs also vary according to the characteristics of the products stored, for example, seasonal items carry a greater risk of perishability, than commodities, such as steel. Firms wish to reduce these costs as much as possible, and the more accurately demand for the products can be predicted, the lower the inventory cost will be, as large inventories will not be necessary. The physical costs of maintaining an inventory is not the only concern, stock outs results in costs as well.

There are four possible outcome of a product being out of stock, consumers may substitute the product for an alternative product, in terms of brand, product, or size (Corstjens and Corstjens, 1999; Hulbert et al., 2003). The customer may defer the purchase until the stock is replenished should it be out of stock, or seek the product at another store. Finally the consumer may even drop the purchase altogether which has the most negative repercussions. For the consumer, the first three outcomes are the most common reactions, and may result in a brand or store switching cost. The brand switching costs relates to the marginal preference of a preferred brand over the next best one. Consumers brand choice behaviour can be regarded as loyal, repertoire, or promiscuous. If brand loyal the customer will always try to buy the same brand, if repertoire, the consumer will readily switch between a collection of brands, and if promiscuous, the consumer is open to any purchase. The level of brand loyalty is also product line dependent with some lines generating greater loyalty. Despite all this, marginal satisfaction can still be brought to zero at the point of purchase through promotions (Corstjens and Corstjens, 1999).

3.2.3. Promotions

Promotion represents the third element in the marketing mix. Keller (1998) describes promotions "as the short term incentive to encourage trial or usage of a product or service". They serve to inform the prospective buyer about the benefits of the product, persuade them to try it and to remind them of the benefits enjoyed through the products use (Berkowitz et al., 2000). There are a number of different forms of promotion including advertising, personal selling, sales promotion, public relations and direct marketing which can be used to influence the consumers' decision process on what goods they purchase. These elements can be used in conjunction to create a promotion strategy to increase the likelihood of a consumer purchasing the product. In developing a promotion strategy it is important to understand the context of the message that the firm wishes to convey, and how the process of communication occurs so that a strategy can be planned and implemented.

3.2.3.1. Communication Process

One of the primary goals of effective promotion is to achieve effective communication. It is "a transactional process between two or more parties whereby meaning is exchanged through the intentional use of symbols" (Engel, 1994 in Holm, 2006). It is the process of conveying a message to others so that is understood accurately. Figure 3.1 shows the communication process.

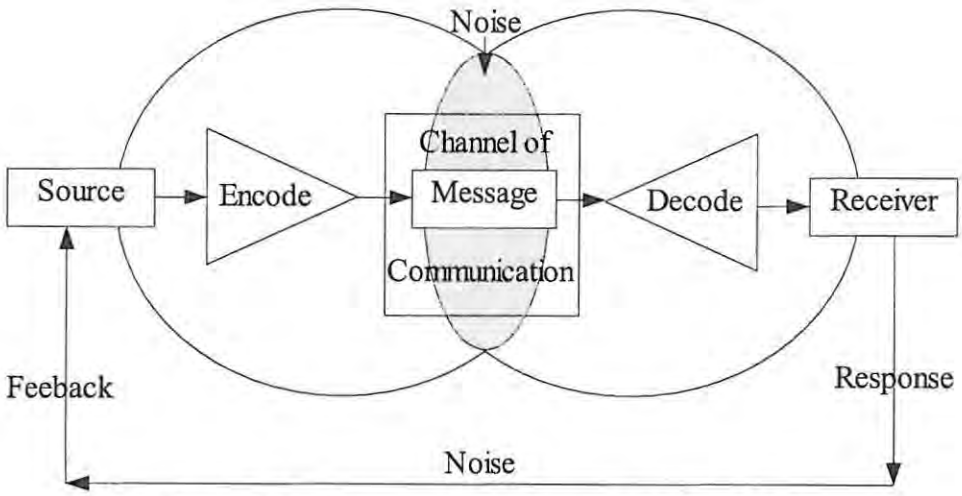


Figure 3.1: Communication Process (Holm, 2006)

The source is the person or company that conveys the message. This information is sent in the form of a message which and can be of any communication medium, such as television, radio or print. The message is sent through a channel of communication, such as an advert. The person who sees, reads or hears the message is the receiver (Holm, 2006).

An understanding of the communication process enable marketers to effectively utilise different promotional mix elements to optimality influence the target audience's buying decisions to purchase specific products. The promotion mix elements discussed below enable the firm to communicate with the consumer either directly or indirectly, and these methods differ in their ability to affect the target audience.

3.2.3.2. Promotional Elements

There are five fundamental methods of communication which firms can use to communicate with consumers. Firms wish to engage in communication with customers to inform them of product features and benefits as well as price and price reductions to encourage them to purchase the products. The communication methods are categorised as either mass or as customised methods. Mass selling involves communicating with groups of consumers simultaneously, whereas customised selling involves an individual interaction between the seller and consumer. Mass communication methods constitute advertising, sales promotions, public relations, and customised include personal selling and direct marketing and are discussed below (Berkowitz et al., 2000).

3.2.3.2.1. Advertising

Advertising is a paid, non personal, mass communication means. Advertising space is bought or paid for and the most common mediums utilised are television, newsprints, or sponsorship (Berkowitz et al., 2000). As advertising is non personal and targeted towards a large audience there is little or no feedback as to how effective it is. As such, market research is particularly important to determine whether the message was understood, and what the most effective medium to reach the target market is for future advertisements. There are a number of advantages to advertising as a means of communication. It is an efficient means of reaching large numbers of people and it also allows the company to control what message it wants to send, as well as to whom it is sent. It controls who the message is sent to through the effective use of advertising mediums as different demographic groups utilise different mediums in their lives. The company can also control how and when the message is sent, including how often as it develops the message, and can control the frequency that the message is distributed (Berkowitz et al., 2000).

Studies conducted in the United States have looked at the power of advertising. A study in 1991-2 showed that approximately half the time advertising "worked". 70% of advertising boosted sales immediately, and the effect was strong in only 30% of the cases (Keller, 1998). Additionally it was discovered that (Keller, 1998) increased sales could come from a single advertisement, blitz campaigns with concentrated exposure are less effective than less frequent long running campaigns. It was also found that advertising was more likely to increase sales and profits than "money off" sales promotions, which often lost money. Although this research is old, it highlights the importance of effective advertising. Advertising is very expensive, and good feedback is very difficult to obtain which are major disadvantages and can be limitations to effectiveness (Berkowitz et al., 2000; Lamb et al., 2002). Sales promotion and advertising are the most frequently used promotional methods in the fast moving consumer goods industry due to their wide market appeal and reach. Other promotional methods also have effect on demand, and need to be at least considered when developing a model to predict it.

3.2.3.2.2. Sales Promotion

Sales promotion is another form of mass promotion, where a short term promise of value is offered to arouse consumer interest in the product. Sales promotions are highly flexible, and are effective at changing short term consumer behaviour. Examples of these include coupons, rebates or samples. The advantages of sales promotions are that it is a short term event which stimulates sales during the event by being an incentive to buy (Mela et al., 1998). The objective of sales promotions is to affect behaviour, rather than attitude (Lamb et al., 2002), and as such sales promotions cannot occur in isolation. Any market gains will be lost once the campaign ends, so it is important to support it with other forms of promotion. In addition, sales promotions cannot occur indefinitely, as this results in promotions losing their effectiveness as the promotions appear normal (Kalwani and Yim, 1992). The simplicity of sales promotions is also a weakness. They can be copied easily by competing firms. Should promotions be copied by firms which sell similar products, a price war may result, as price remains one of the few remaining factors that firms can differentiate themselves with. Promotions are also open to abuse from unethical businesses, as these firms can claim coupon rebates for all products sold, even if the customer did not claim the rebate (Berkowitz et al., 2000).

There are a number of alternative sales promotion techniques and some of the more common of these are discussed below. Coupons offer a discounted price to the consumer in the form of a voucher offering a price reduction. The coupon is an effective method of price discrimination, as the coupon redemption rate is different for consumers with different price sensitivities (Levedahl, 1984; Narasimhan C, 1984 in Son et al., 2006). The marginal price for loyal customers is unaffected, thus their purchasing behaviour is unaffected, and new customers are attracted which have a lower price elasticity, and are willing to switch brands (Levedahl, 1984). Offering coupons increases units sold immediately upon distribution of the coupons as new customers are attracted, but also lowers revenues from already loyal customers who make the same purchases at a lower rate (Leone and Srinivasan, 1996; Berkowitz et al., 2000). Coupons may also adjust the purchasing behaviour of consumers depending on the frequency of the offering. If the coupon offer is frequent, consumers may adjust the number of units they buy in anticipation of a future offering (Levedahl, 1984).

Deals are short term price reductions. They are commonly used to encourage trial among potential users, or as a reaction to a competitors actions. These responses are commonly an effort to prevent the consumer from using the competitors product, for example, if a competitor produces a new product, the company may respond with a two for the price of one deal, thus building a supply in the consumers home, and as a result preventing their need to restock. This makes the new products' introduction significantly more difficult (Morris, 1989; Lamb et al., 2002; Berkowitz et al., 2000).

Premiums are when a product is offered free, or at a significant discount over the retail price. For example, a free toy with a certain children's meal at a restaurant, or a R100 discount on something on the production of three vouchers. This is advantageous as consumers are attracted to free products, and this builds goodwill and they may continue to buy once the promotion is over. However, a risk of premiums is consumers may buy the product during the promotion purely for the freebie, and not migrate to it once the offer is finished (Berkowitz et al., 2000; Lamb et al., 2002).

Contests are where consumers apply a skill or creative thinking to win a prize. The prize is often very appealing, and it encourages consumers to purchase that product. The disadvantage is that many customers will only migrate to the product for the duration of the competition, or may be unwilling to enter the competition as it requires some form of effort of skill (Morris, 1989; Berkowitz et al., 2000; Lamb et al., 2002). Alternatively to contest based promotion, sweepstakes can be held where the prize is awarded on a lottery basis. It has similar advantages in encouraging migration to the product for the duration of the promotion, but does not necessarily mean that the product is the consumers preferred product, so sales often drop after a sweepstakes competition as consumers migrate to their original product preferences (Berkowitz et al., 2000; Lamb et al., 2002).

In an effort to get consumers to test the product, companies offer samples which are free or at a reduced price sample. Commonly a single use package is used, and offered free as a trial.. As it is free consumers are very likely to try it, and it is hoped that if consumers like the sample, they will remember it, and purchase it in the future. Although this is very likely to get consumers to try the product, it is a very expensive method of promotion for the company (Berkowitz et al., 2000).

It is often said that it cost to obtain new clients than to retain existing ones. As a result continuity programs are a promotion tool used to encourage and reward repeat purchases by loyal customers. The most common example of this is airline frequent flier cards. It successfully encourages repeat purchases and loyalty, thus ensuring a stable market (Son et al., 2006), but then benefits that are rewarded to loyal customers are expensive for the firm (Berkowitz et al., 2000).

Research suggests that up of 70% of the consumer buying decisions are made in the store (Keller, 1998), so it is essential to attract the consumers attention at the point the decision is made. As a result in store promotions are a commonly used form of promotion using point of purchase displays. Point of purchase promotions take the form of advertising signs which display or highlight the product including shelf labelling, signs, or gondola ends (end of isle promotion stands). Point of purchase promotions also support and act as a reminder for other forms of promotion that may be in progress, and provides product visibility to the customer. The major disadvantage of in store promotions are the difficulties, and expense in getting the retailer to allow the promotion materials to be placed in high shopper traffic space as there is much competition for prime space between firms (Berkowitz et al., 2000).

Although sales promotions appear to have a positive effect in attracting new customers, Mela et al. (1998) argue that price promotions primarily impact brand distinctiveness, and that deals decrease the loyalty and increase price sensitivity of the promoting brand more than they affect other brands (Erdem et al., 2002). In effect, discounting (promotions) may have negative long-term consequences for brands even though promotions can induce dramatic sales increases in the short-term (Gedenk and Neslin, 1997 in Mela et al., 1998). However, these short term impacts must be understood to determine the effects in demand that are induced by specific promotions.

3.2.3.2.3. Public Relations

The final mass communication method is public relations and seeks to influence the feelings, opinions or beliefs held by consumers, stockholders, employers, and any other public stakeholders. Tools such as annual reports and image management can be used. Publicity is also used, which is a non personal, indirectly paid presentation of the organisation, goods or services that are being offered. It can take the form of a news article, editorial, or product announcement. The advantage of public relations is that it appears more credible in the consumers mind, as it does not originate from the company. The major disadvantage is that it is very difficult to get the media's cooperation, and it also cannot be guaranteed what will be broadcast (Berkowitz et al., 2000; Lamb et al., 2002).

3.2.3.2.4. Personal Selling

Personal selling is a form of customised marketing that encourages a two way flow of information between a buyer and a seller (Lamb et al., 2002). It is designed to influence a person's or group's decision through face to face communication between the buyer and the seller. More commonly in the digital era, this is occurring through customised internet interfaces. Personal selling has significant advantages in that it is more persuasive, as it is conducted on a face to face basis, and that the target audience can be carefully selected. There is also immediate feedback, and there is a learning of complex information which can be used to improve the process of personal selling. This flexibility of personal selling can also be seen as a disadvantage, as the message can be adjusted between sales people, so that there is no consistent message. In addition, it is exceptionally expensive on a per person basis (Berkowitz et al., 2000).

3.2.3.2.5. Direct Marketing

Direct Marketing is another form of customised marketing. "The objective is to communicate with a customer from one person to another in an environment that is not a permanent retail location" (Duffy, 2005) with the objective of generating a response in the form of an order, a request for additional information, or a visit to a retail outlet. Direct marketing comes in a number of forms, face to face selling, direct mail, or catalogues. Like personal selling, messages are easy to adapt to a specific audience. However, most forms of direct marketing require an up to date database which is expensive to develop and maintain. With increasing concerns about privacy, there has been a decrease in response rates among consumers. Successful companies realise this and combine direct marketing with other forms of promotion (Berkowitz et al., 2000; Lamb et al., 2002).

3.2.3.2.6. Promotion Conclusion

Research has strongly suggested that advertising has a long term effect on sales, and promotion a short term one (Takada and Bass, 1998). Marketing communication must be based on a totally integrated plan, balancing all forms of promotion, and with the appropriate product positioning to maximise market reach and influence to obtain the desired consumer response (Gruenwald, 1997). A comprehensive understanding of the promotional strategies enables greater accuracy in demand planning. Despite a perfect promotional strategy, a desired product is still necessary for the marketing plan to be a success.

3.2.4. Product

A product is a good, service, or idea consisting of a bundle of tangible and intangible attributes that satisfy consumers and is received in exchange for money or a unit of value (Berkowitz et al., 2000). It is seen as the basic resource involved in the exchange process (Hakansson and Waluszewski, 2005). Products are classified in two main ways. They are grouped as product lines with other products satisfying closely related customer needs or they are classified by type of user, or degree of tangibility.

A product line is a group of products that are closely related because they satisfy needs that are closely related (Ferrell and Hartline, 2005). They are also sold to the same consumer group. The product line considered in this thesis is the deodorant product line. Within a product line, each individual item is specific by its brand, size, unique feature and price. The benefit of multiple product lines for a single producer are economies of scale in production and raw material purchasing. Often product lines use similar components in production, which reduces inventory handling costs, and there are sales and distribution efficiencies, as a full line can be offered to retailers as opposed to individual products (Ferrell and Hartline, 2005).

Products are classified to aid in developing marketing strategies for similar products. Products are typically classified by the degree of tangibility, or by the type of user. When classified by tangibility, tangible products can be described as non-durable where they are used in a single or few uses, for example food. Another classifications is durable where products last over a large number of uses, for example, audio equipment. Intangible products are described as services which are activities, benefits or satisfaction that are for sale (Ferrell and Hartline, 2005).

The other form of classification is classification by type of user. Primary consumers purchase industrial goods, which are used for production of other goods, which are then sold to the final consumer, whilst consumer goods are purchased by the final user (Ferrell and Hartline, 2005). While a consumer may be satisfied with their current purchases and range of choices, their demands are constantly changing and as such new products are constantly being developed to meet these needs.

New products are constantly being introduced to the market. These products are supported by research, which suggests that consumers have a need for the product, and would purchase it. New products can lift the morale of sales staff, excite customers, as well as steal the advertising spotlight (Magrath, 1995). Although new products offer benefits to the consumer apart from additional features and choice, it also increases competition between firms, which mean that firms may find it difficult to retain their existing market shares. In addition to additional competitors, firms should be aware of the change in consumer tastes and competitor actions which may affect the product and its success in the market. This is known as the product's life cycle (Berkowitz et al., 2000).

3.2.4.1. Product Life Cycle

The product life cycle (PLC) explains the stages that a product goes through in the marketplace, from introduction to decline. The PLC is also useful to explain and predict other actions within the marketing mix. The PLC is illustrated in figure 3.2 below (Ferrell and Hartline, 2005) and consists of four stages; introduction, growth, maturity and decline.

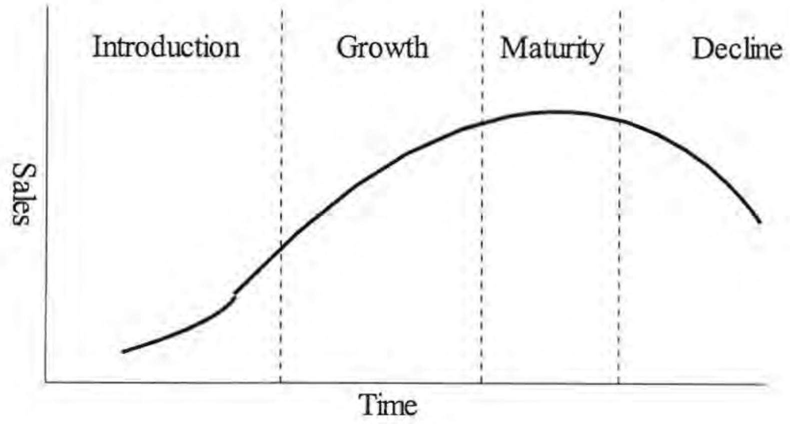


Figure 3.2: Product Life Cycle
(Berkowitz et al., 2000; Ferrell and Hartline, 2005)

The introduction stage is when the product is first introduced into the market, sales increase slowly and profits are minimal. Profits are low as research costs are often substantial, and need to be recouped first. The major objectives are to create consumer awareness and encourage consumers to try the product, as well as to educate them on its use (Berkowitz et al., 2000; Ferrell and Hartline, 2005). These objectives lead to a distinctive marketing mix strategy. There is often massive promotion expenditure to stimulate demand (Catarelli, 1980). The expenditure is not to create a brand, as often a new product is the only one in the category, so there is no competition but rather to create consumer awareness. Pricing is either low or high, in the form of price skimming or penetration, with the purpose of harnessing maximum profits, or maximise volume and to discourage competitors from entering the market.

The growth stage is characterised by rapid increases in sales (Ferrell and Hartline, 2005). The number of competitors as well as users increases, and as a result sales increase. There are an increasing number of first time as well as repeat users. With the increased competition, there is a need to differentiate the product from competitors, so a result, improved versions, or additional features are added. Promotion changes from an informative role in the introductory stage, to where it highlights the competitive differences. In addition to the product variants, distribution improves to maximise customer exposure, and increase the chances of a sale (Berkowitz et al., 2000). There are two main objectives; to gain market share whilst competing against competitors, and to repay investments to ensure long term product viability (Ferrell and Hartline, 2005).

The maturity stage is the longest stage (Ferrell and Hartline, 2005) and is typical of a slowing of overall industry sales, as the marginal competitors exit the industry and most consumers are repeat buyers. Sales therefore increase at a declining rate, or remain fairly constant. The marketing mix objectives change to ensure that the market share is maintained (Hart and Tzokas, 2000). Profit decreases as there is significant competition between the existing players, and the cost of attracting new consumers increases significantly. Promotion is done to serve as a reminder of the product, with the intent of maintaining market share, and to promote additional product differentiation which may have occurred to attract more niche customers. The products are also priced in line with competitors as to defend the market share (Ferrell and Hartline, 2005).

Sales and profits eventually being to drop and this is referred to as the decline stage. This typically occurs due to uncontrollable external environmental factors, such as new technologies, or product replacements. When a product enters the decline stage, the company assumes two strategies, deletion or harvesting (Berkowitz et al., 2000; Ferrell and Hartline, 2005). Deletion occurs when the product line is dropped completely. As there are still consumers who use the product, this is difficult as loyal customers may become dissatisfied and may not migrate to new products. Harvesting is when the company maintains the product, but reduces marketing. As a result the product is continued to be offered, but no time or money is spent on it. The purpose of this is to meet consumer needs, while costing the firm very little (Berkowitz et al., 2000). In order to prevent the decline stage and extend the profitable stages of the PLC, successful management of the PLC is necessary.

3.2.4.2. Product Life Cycle Management

Product life cycle management is the process of managing the product through its successive stages in their life cycle (Berkowitz et al., 2000). It is important in extending the life of the product, in new product innovations and getting the product to market faster as the industry and market evolves (Kitchen, 1989; Sudarsan et al., 2005). The product must be made to exist for as long as possible, so as to maximise total profit through the product's life.

PLC management involves the need to balance fast responses to changing consumer demands with competitive pressure to seek cost reductions in sourcing, manufacturing and distribution Jüttner et al., 2006). There are three ways of doing this: product modification, market modifications, and repositioning the product. If the product life cycle is successfully managed it will take on the appearance as shown in figure 3.3.

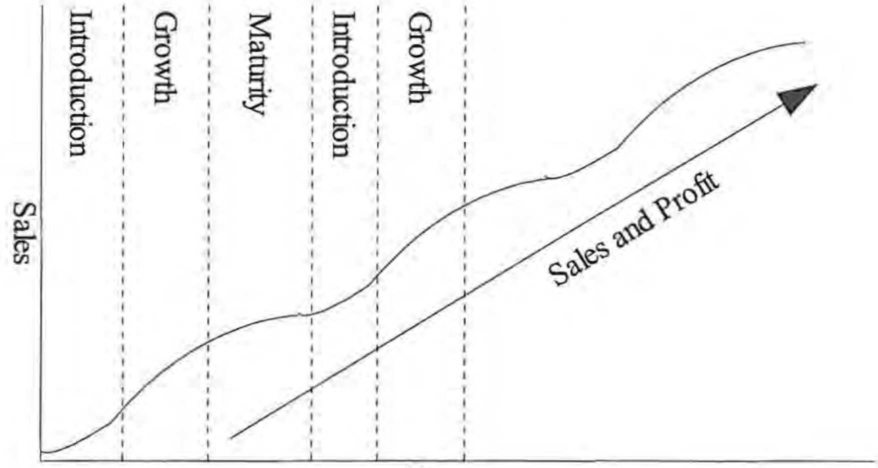


Figure 3.3: Managed Product Life Cycle

Product modification involves adjusting the products characteristics, such as quality, appearance, or performance to enhance the products sales. New features, such as scent or even packaging, can be used to give the impression of a revised product (Berkowitz et al., 2000).

Market modification involves attempting to increase the market's use of the product among existing customers through the creation of new use situations. It also involves finding new customers. An example of attempting to increase the product usage may be by marketing it differently, such as marketing fruit juice as a healthy all day drink as opposed to a breakfast drink. New situations can be created by marketing an alternative use for customers, for example, table salt as a stain remover, as well as a food flavourant. Finding new users can be done by creating a variant for a new market, such as a video game for younger children as opposed to the traditional teenage market (Berkowitz et al., 2000).

The product can be repositioned by changing the consumers' impression of a product. For example 4x4 vehicles have been repositioned from exclusive off road vehicles to utility city vehicles as well. There are numerous reasons to attempt this, including reacting to a consumer's position which is corroding market share. Products can be repositioned to reach new markets. A rising trend offers an opportunity, and repositioning an existing product to catch this is a good growth strategy. Finally a product can be repositioned by trading the product up or down the product, by increasing the quality, attracting a different market, or by trading down, and lowering the quality and features available to attract a different market (Berkowitz et al., 2000). In the effort to be attractive to additional markets and maintain traditional markets, firms may enter into damaging competitive situations.

3.3. Mature Markets and Price Wars

A mature market is one where there has been growth, and demand has become constant, firms lose their competitive advantage as there are no further technological innovations, and there are many market players. In such an industry, the competitors must fight to win sales (Magrath, 1995). A structured, hierarchical approach using the key drivers of the marketing mix has been suggested as a method to gain market share (Weber and Dholakia, 1998), but this does not guarantee success in a mature market.

A mature market is typically characterised by two phenomena: overcapacity, and product parity. During growth periods, capacity is expanded ahead of competitors, but when the demand slows, the excess capacity cannot be stopped in time, and as a result capital is tied up in equipment, which remains unused. Product parity occurs when there are very few advances given by technological innovation, and to the consumer all products are seen as equal, and customers can swap easily between products, and as a result it is very difficult to maintain repeat customers as they can change product based purely on price (Corstjens and Corstjens, 1999).

Market competition is when there is an implicit agreement between producers to target marginally different consumers needs (Corstjens and Corstjens, 1999). This results in clear segmentation and companies being able to better serve specific needs. As a result the total value of the market increases as there is less direct competition. To attract the specific markets, promotion is done via advertising, to alert consumers of the actual product, and how it would meet their specific needs. For the greatest success in the market orientated industry, the consumers and their buying decisions must be understood, which should be well defined in the marketing strategy and marketing mix (Corstjens and Corstjens, 1999).

3.4. Conclusion

Successful marketing requires the successful meshing of a number of different organisational facets. The marketing mix and Kotlers 4P's of product, price, place and promotion provide excellent insight into the key components of marketing. However, it is a successful marketing program that will integrate them to provide maximum sales and value for the firm.

Firms aim to develop a product that the market desires, and then balancing the price according to the firms cost structure and profit objectives. An appropriate promotional strategy is required to ensure the products benefits are communicated successfully to the correct target market. Sufficient logistics are required to ensure that the product is at the right place while minimising inventory costs.

A thorough understanding of how the elements of the marketing mix affect the price is important in the process of estimating demand. The point in the product life cycle determines the initial pricing strategy, which adjusts as the product moves through the life cycle which is further affected by the product's elasticity. The effect, and length of the effect of promotions is also important, as certain forms of promotion result in immediate short term demand, whereas others result in longer term product loyalty and demand increases. Assuming that the product is positioned correctly for the target market, and the correct locations determined for sales, they have a constant effect on sales, and are not variable factors in the changing levels demand. The successful determination of this is a complex task but done successfully, firms are rewarded with strong profits. Firms are further able to reduce costs through the accurate prediction of sales lowering inventory costs which is aided by the thorough understanding of the affects of the marketing mix on the consumer.

In estimating demand, firms can effectively utilise all this knowledge of the influence that the marketing mix has on sales, and then be able to estimate the sales quantity for the next period, and the supply chain can then effectively control the inventory to minimise costs. In addition, the production department can make the scheduling and arrange the facility utilisation. Such an action may cause the production cost to decrease (Kuo and Xue, 1998). Therefore, using all the information obtained through the marketing process to obtain an accurate forecast is crucial for firms. This thesis aims to utilise this knowledge in understanding the factors which influence consumer demand, and from this understanding develop a model which accurately predicts demand.

CHAPTER 4

TIME SERIES

4.1. Introduction

Good forecasts are vital in many areas of scientific, industrial, symmetrical and economic activities. This chapter discusses time-series which is a collection of observations on a variable *y*, taken at intervals over time (Harvey, 1994). Forecasts are made on the basis of data comprising one or more times-series (Chatfield, 2000) and is extrapolated to make predictions. This chapter explains the basic fundamentals of time series, and describes a number of univariate models. Various forecasting methods within the time series model, such as smoothing methods and the ARIMA method, are also explored.

4.2. Stochastic Processes

Time series is primarily concerned with modelling data around a stochastic process, which are described as a "collection of random variables which are ordered around time" (Harvey, 1994:9), where the time can be either continuous or discrete. The moments of a stochastic process are defined with respect to the distribution of the random variables $X_{t_1}, \ldots, X_{t_T}$. The mean at time t is

$$\mu_t = E(X_t) = \int x\, f(x,t)dx, \qquad t=1,\ldots,T \tag{4.1}$$

where $f(x,t)$ denotes the density function of X at time t. This can be interpreted as the average value or mean of X_T taken over all possible realisations. The second moment, or variance at time t is defined by (Cryer, 1986)

$$Var(X_t) = E[(X_t-\mu_t)^2], \qquad t=1,\ldots,T \tag{4.2}$$

and the covariance between X_t and X_{t-k} is given by

$$Cov(X_t, X_{t-k}) = E[(X_t-\mu_t)(X_{t-k}-\mu_{t-k})], \qquad t=1,\ldots,T. \tag{4.3}$$

If a model can be found which approximates the random behaviour of a stochastic process, forecasts can be computed (Chatfield, 2000) and this requires the stochastic process to be stationary, or the properties of the underlying model do not change over time. A model is first order stationary when $E(X_t)=\mu$, $Var(X_t)=\sigma_0$, and $Cov(X_t, X_{t-k})=\sigma_k$ are constant, i.e. independent of time (Johnston and DiNardo, 1997). The simplest example of a covariance stationary stochastic process is a sequence of uncorrelated random variables with constant mean and variance. This is called White Noise, where (Harvey, 1994)

$$E(X_t X_{t+k}) = 0 \quad if\ k \neq 0 \\ = \sigma^2 \quad if\ k = 0 \tag{4.4}$$

The covariance between X_t and X_{t+k} for different values of t and k is called autocovariance, which depends on lag k, such that

$$Cov[X_t, X_{t+k}] = \sigma_k \tag{4.5}$$

The set of autocovariance coefficients, $\{\sigma_k\}$, for $k = 0,1,2,\ldots,$ constitutes the autocovariance function. If the function is strictly stationary, the autocorrelation becomes (Cryer, 1986)

$$\rho_k = \frac{\sigma_k}{\sigma_0}. \tag{4.6}$$

In a stationary process ρ_k measures the correlation at lag k between X_t and X_{t+k}. This is an even function where $\rho_k = \rho_{-k}$, and $|\rho_k| \leqslant 1$.

The correlogram is a very useful tool in time series analysis, as it can be used to provide an estimate of the autocorrelation function, as well as "assessing the behaviour and properties of a time series" (Chatfield, 2000:31). The sample covariance at lag k is calculated by

$$c_k = \frac{1}{n} \sum_{t=1}^{N-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \tag{4.7}$$

for $k = 0,1,2,\ldots$, and the sample autocorrelation at lag k is calculated by (Box and Jenkins, 1968)

$$r_k = \frac{c_k}{c_0}. \tag{4.8}$$

This is used to develop a correlogram which appears as a plot of r_k against lag k. For data from a stationary process it can be shown that the correlogram provides an estimate of the theoretical autocorrelation function defined in (4.2). It follows that if the series is nonstationary, the correlogram does not provide an estimate of anything. In those cases, the values of the correlogram do not converge to zero, and the only value obtained is in the indication that the series is not stationary (Chatfield, 2000).

The interpretation of a correlogram is very difficult, especially where the sample size n<100, which results in the autocorrelations having large variances. Common patterns for a stationary time series is to see short term correlation where the first three of four r_k are significantly different from zero.

For seasonal series there is likely to be a large positive value of r_k at the seasonal period, which may still be present after the seasonal pattern has been removed (Chatfield, 2000). Thus the correlogram can easily be used to see if seasonality is present. For a series with a trend, the correlogram does not decrease to zero until a high lag has been reached, which suggests that some form of trend removal is necessary to make the time series stationary known as transformations (Cryer, 1986; Chatfield, 2000).

4.3. Transformations

The creation of effective time series models requires a stationary time series. However, in practise data are seldom stationary (Taio, 2001). In Chapter 2 various trends were described, such as growth trends, seasonal trends, other cyclic variations or irregular fluctuations (Lewis, 1975; Chatfield, 2000).

It is important that these trends are removed to create a stationary time series. There are a number of methods of achieving this, differencing for the removal of trends, or first order nonstationarity, and through transformations to remove second order nonstationarity (Mills, 1990).

4.3.1. Trend Removal

It is assumed that a time series can be written as $X_t=\mu_t+\varepsilon_t$ where μ_t represents the trend, thus $\varepsilon_t=X_t-\mu_t$, which is the error after the removal of the trend (Mills, 1990). Trends can be tested for in a time series through a 'runs test' where residual randomness is tested for. Where too few runs is an indication of a trend, and too many is an indication of a periodic component (Cryer, 1986). Additionally if the series is split into two equal parts, of $X_1,\ldots,X_{T_1}$ and $X_{T_1+1},\ldots,X_T$ of equal length, the means of these groups can be tested, using a pooled t-test if the data are normally and randomly distributed, or by using a Mann-Whitney test if it is not normally distributed. If H_0 states that the means are equal and is rejected, there is an indication of the presence of a trend.

The most common method of removing a linear or other p^{th} order polynomial trend is through differencing, where a past value of a variable is subtracted from its current value (Cromwell et al., 1994). Using the transformed variable, $\Delta X_t=X_t-X_{t-1}$ is sufficient for removing a linear trend, where the differencing is repeated until the mean sum of square errors is a minimum, and the next difference's mean is almost zero. When stationarity is obtained, the correlogram of the differences will show rapidly decreasing autocorrelations. The lag operation, L is a useful way of describing the differencing operation, the simplest operation is $LX_t = X_{t-1}$, and higher powers of L result in longer lags, $L^2 X_t = X_{t-2}$ and so on. These difference operations act as a filter to remove linear and polynomial trend components (Harvey, 1994).

4.3.2. Removing Second Order Nonstationarity

For time series which display a non constant variance it is important that this second order nonstationarity is stabilised using transformations (Chatfield, 1989). A transformation may also be necessary if the data do not appear to be normally distributed or are skewed (Mills, 1990).

There are a number of different transformations which are useful in difference scenarios. The transformation of $\sqrt{(X_t)}$ is used when the variance of X_t is proportional to $E(X_t)$. It is most commonly used when the $X_t's$ are counts from a Poisson distribution (Cryer, 1986). The $\ln X_t$ transformation is very commonly used, especially amongst economists, as the log change in the variates estimate the percentage of change or rate of return. It is especially useful when the $X_t's$ vary from one to thousands, but all X_t are positive (Chatfield, 2000). Box-Cox transformations are essentially power transformations where the time series x_t is transformed by

$$y_t = \begin{cases} (X_t^{\lambda} - 1)/\lambda & \lambda \neq 0 \\ \log X_t & \lambda = 0 \end{cases} \tag{4.9}$$

where λ denotes the transformation parameter which can be estimated through a formal statistical procedure, such as the maximum likelihood method (Chatfield, 1989) and is typically $0 \leqslant \lambda \leqslant 1$. The Box-Cox transformation makes y_t a continuous function which helps to reduce asymmetry (Mills, 1990). There are a number of different version of the Box-Cox transformations to account for negative values and skew transformations (Sakia, 1992). The estimation of λ has been described using maximum likelihood methods, Bayesian methods, and later computer methods (Sakia, 1992).

"Transformations are widely used, but there is little evidence which suggests that the use of non-linear transformations improves forecasts" (Chatfield, 2000) and there can be difficulties in interpreting a model for a transformed series. Box and Jenkins maintain that this is because the incorrect type of transformation was applied (Makridakis, 1978). In addition when forecasts are transformed back to the original scale of measurement they will often be biases if the transformed forecast was unbiased. However, transformations are useful when there is asymmetry in the variance, or there are severe changes in the variance (Chatfield, 2000). Once a time series is stationary, there are a number of techniques which can be used to model the series, which are later extrapolated for forecasting.

A unit root test can be used to determine whether the time series is nonstationary and requires differencing or another form of trend removal to make forecasts. In a unit root test a regression of the time series is estimated, where

$$\Delta X_t = a_1 + a_2 t + a_3 X_{t-1} + \epsilon_t \tag{4.10}$$

and X_{t-1} is a one period lag of X, t is the trend variable, and a_i are coefficients. The null hypothesis states that the coefficient of X_{t-1} is zero, or nonstationary, which is called the unit root hypothesis. Using the Dickey Fuller test, if the τ value of a_3 is greater (in absolute value) than the Dickey Fuller critical value, the unit root is rejected, and can conclude that the series is stationary (Gujarati, 2006). Diebold and Kilian (1999) argue that the unit test is also useful for pre-testing, and that it improves forecast accuracy over routinely differencing the data.

4.4. Univariate Time Series Classes

Forecasting techniques are commonly based on a time series model. Thus familiarity of the types of models is important, as it allows a better understanding of what processes could be used in the forecasting process. The most common form of model, is the univariate model which "describes the distribution of a single random variable at time *t*, namely X_t, in terms of its relationship with past value of X_t and its relationship with white noise shocks" (Chatfield, 2000:85). There are seven common univariate models which can be used to generate more complicated ones for use in forecasting.

4.4.1. White Noise

A White noise process is a purely random process which is a sequence of uncorrelated, identically distributed random variables, with zero mean, and a constant variance. This process is stationary, with an autocorrelation given as (Maddala, 2003)

$$\rho_k=\begin{cases}1 & if\ k=0\\ 0 & otherwise\end{cases} \tag{4.11}$$

This model is also known as "uncorrelated white noise", or the "error process". It is rarely used to describe data, but rather to model random disturbances in complicated processes (Chatfield, 2000). ϵ_t is used in this thesis to denote a purely random precess with zero mean and variance of σ^2.

4.4.2. Random Walk

The random walk is a simplest nonstationary model, and describes a time series as (Harvey, 1994)

$$X_t=X_{t-1}+\epsilon_t \tag{4.12}$$

This model is most commonly used in finance, where the price of a share on any particular day is equal to the price on the previous day plus a random change in the share price, either positive or negative. The random change generally cannot be forecasted as it has properties similar to a purely random process. It does not necessarily form a stationary process as the variance may increase over time. However, the first difference, i.e. X_t-X_{t-1} is stationary (Chatfield, 2000).

4.4.3. Autoregressive Processes

Yule first introduced the Autoregressive Moving Average (ARMA) models in 1927 when he stated that time can be regarded as a stochastic process (De Gooijer and Hyndman, 2006). The simplest being an autoregressive process (AR) which takes on the appearance of a regression model, but X_t is regressed on past values of X_t. It has the form of (Shumway, 1988; Chatfield, 1989)

$$X_t=\phi_1 X_{t-1}+\phi_2 X_{t-2}+...+\phi_p X_{t-p}+\epsilon_t \tag{4.13}$$

where ϕ_i are the parameters. The simplest case is the first order process of AR(1), given as $X_t=\phi X_{t-1}+\epsilon_t$. If $\phi=1$, the model reduces to a random walk as in (4.8) when the model is nonstationary. If $|\phi|>1$, the series is explosively nonstationary and if $|\phi|<1$, then the process is stationary with autocorrelation function $\rho_k=\phi^k$ for $k=1,2,...$ (Box and Jenkins, 1968; Chatfield, 2000).

4.4.4. Moving Average Processes

A Moving Average (MA) is a series where X_t is dependent on ϵ_t as well as q previous random shocks. As such it is given by a weighted linear sum of the last q random shocks

$$X_t=\epsilon_t-\theta^1\epsilon_{t-1}-\theta^2\epsilon_{t-2}-...-\theta^q\epsilon_{t-q} \tag{4.14}$$

where ϵ_t denotes a purely random process with zero mean and constant variance σ^2 and θ^i are the parameters. The simplest case is a first order moving average model, or MA(1), which is given by $X_t = \epsilon_t - \theta\epsilon_{t-1}$. This model is stationary for all values of θ, with the autocorrelation function given as (Box and Jenkins, 1968; Chatfield, 1989)

$$\rho_k = \begin{cases} 1 & for\, k=0 \\ \dfrac{-\theta}{(1+\theta^2)} & for\, k=1 \\ 0 & for\, k>1. \end{cases} \tag{4.15}$$

Thus the autocorrelation cuts off after 1 lag. Although there are no restrictions for an MA process to be stationary, it is generally necessary to impose conditions on θ^i to ensure that the process satisfies a condition called invertibility. An MA(1) model of $X_t = \epsilon_t - \theta\epsilon_{t-1}$ may be inverted to give $\epsilon_t = X_t + \theta^1 X_{t-1} + \theta^2 X_{t-2}$. If $|\theta| < 1$ this substitution can be continued into the past such that (Chatfield, 2000)

$$X_t = (-\theta^1 X_{t-1} - \theta^2 X_{t-2} - \ldots) + \epsilon_t. \tag{4.16}$$

As such, if $|\theta| < 1$ then the MA(1) model can be inverted into an infinite order autoregressive process, thus MA(1) is invertible. Similarly for an MA(q) model with polynomial equation

$$\theta(x) = 1 - \theta^1 x - \theta^2 x_2 - \ldots - \theta^q x_q. \tag{4.17}$$

is invertible (Cryer, 1986).

The backshift operator is commonly used in the notation of MA and AR models, or models with an MA or AR derivative. It is denoted by B [7], and shifts time back 1 unit to form a new series, i.e. $B(X_t) = X_{t-1}$ (Shumway, 1988; Kendall and Ord, 1990). The backshift operator is linear as for any constants, *a*, *b*, and *c*, and series X_t and ϵ_t, it is clear that

$$B(aX_t + b\epsilon_t + c) = a[B(X_t)] + b[B(\epsilon_t)] + c. \tag{4.18}$$

As such the MA(1) model can be written as

$$\begin{aligned} X_t &= \epsilon_t - \theta\epsilon_{t-1} \\ &= \epsilon_t - \theta B(\epsilon_t) \\ &= (1-\theta B)(\epsilon_t) \\ &= \theta(B)\epsilon_t \end{aligned} \tag{4.19}$$

7 B is also commonly used as the notation for the lag operator L which is most commonly used by economists (Chatfield, 2000)

where $\theta(B)$ is the MA characteristic. Similarly for an MA(q) model, which can be written as $X_t=(1-\theta^1 B-\theta^2 B^2-...-\theta^q B^q)\epsilon_t$ or $X_t=\theta(B)\epsilon_t$. The backshift operator can also be applied to autoregressive models, where all the X_t terms all moved to the left hand side, such that an AR(q) model takes the form of $X_t-\phi_1 X_{t-1}-\phi_2 X_{t-2}-...-\phi_q X_{t-q}=\epsilon_t$, or $\phi(B)X_t=\epsilon_t$. The AR model is also invertible, of the form $X_t=\Psi(B)\epsilon_t$ where $\Psi(B)$ is the AR characteristic (Cryer, 1986; Chatfield, 2000).

4.4.5. Mixed ARMA Models

In practise, there are commonly too many parameters to estimate from a finite data set, thus a model is necessary that estimates a model with as few parameters as possible. A mixed autoregressive moving average model with p autoregressive terms and q moving average terms is abbreviated as ARMA(p,q), and is written as (Chatfield, 1989)

$$X_t=\phi_1 X_{t-1}+...+\phi_p X_{t-p}+\epsilon_t-\theta^1\epsilon_{t-1}-...-\theta^q\epsilon_{t-q}. \tag{4.20}$$

Using the backshift operator it is written as (Chatfield, 2000)

$$\Phi(B)X_t=\Theta(B)\epsilon_t \tag{4.21}$$

where $\phi(B)$ and $\theta(B)$ are polynomials of finite order p and q. The AR process has values of α_i which makes the process stationary such that the roots of $\phi(B)=0$ lie outside the unit circle of $(-1,1)$. For the MA process the values of β_i which make the process invertible are such that the roots of $\theta(B)=0$ lie outside the unit circle (Chatfield, 1989). In reality this model is not known, and as such has to be estimated from the data (Taio, 2001).

4.4.6. ARIMA Process

As most time series in practise are nonstationary, we cannot apply stationary AR, MA, or ARMA processes directly (Chatfield, 1989). As discussed previously, differencing can be used to make nonstationary processes stationary. The first difference would be $(X_t-X_{t-1})=(1-B)X_t$ and differenced repeatedly such that the d^{th} difference may be written as $(1-B)^d X_t$ (Box and Jenkins, 1968). If the original series is differenced d times before fitting an ARMA(p,q) process, then the model for the undifferenced series is said to be an ARIMA(p,d,q), where 'I' stands for integrated and d denotes the number of differences. This is defined as (Chatfield, 1989; Enders, 2004)

$$\phi(B)(1-B)^d X_t=\theta(B)\epsilon_t. \tag{4.22}$$

If B is replaced by X, the AR process has d roots on the unit circle, as where $x=1$, indicating that the process is nonstationary, which is why differencing is necessary. When $\phi(B)$ and $\theta(B)$ are both equal to unity, i.e p and q are both zero, the model reduces to an ARIMA(0,1,0) model, which is given by $X_t - X_{t-1} = \epsilon_t$, which is the same as the random walk model in (4.4.2). The main difficulty with fitting AR and MA models is assessing the order of the process p and q, rather than the estimation of ϕ and θ (Chatfield, 1989). With ARIMA models there is an additional difficulty is assessing the order of differencing. The unit root test is appropriate for testing how many differences are necessary. However, most commonly the series is differenced until the correlogram decreases to zero quickly. Once the model has been made stationary through differencing, an ARMA model can be fitted (Chatfield, 2000).

4.4.7. SARIMA Process

If the time series is seasonal, with time periods per year, quarterly or monthly then a seasonal ARIMA model may be used, which is known as a SARIMA process. Seasonal differences can be denoted as

$$X_t - X_{t-s} = (1 - B^s) X_t. \tag{4.23}$$

The seasonal autoregressive term is where one X_t depends linearly on X_{t-s}. A SARIMA model with non-seasonal terms (*p*,*d*,q) and seasonal terms of order (P,D,Q) is described as a SARIMA(p,d,q) x $(P,D,Q)_s$ model and can be written as

$$\phi(B)\Phi(B^s)(1-B)^d (1-B^s)^D X_t = \theta(B)\Theta(B^s)\epsilon_t \tag{4.24}$$

where Φ, Θ denote polynomials in B^S of order P, Q respectively. For example a model with 1 difference, and 1 MA process with monthly data is shown as

$$(1-B)(1-B^{12})^D X_t = (1+\theta B)(1+\Theta B^{12})\epsilon_t. \tag{4.25}$$

When selecting a SARIMA model appropriate values of differencing must be selected for both seasonal (D) and non-seasonal (d) to make the model stationary and remove most of the seasonality. Then an ARMA model can be fitted (Chatfield, 2000).

4.5. Multivariate Time Series

Having successfully defined a number of the more common univariate technique a cursory mention of multivariate time series is made. It was first developed by Quenoille in 1957, but little attention was paid to it until computing power and access to computers improved in the 1970s (De Gooijer and Hyndman, 2006). Observations are often taken over two or more time series simultaneously, it is necessary to describe the relationship between the series, and then use this model to make forecasts. This is done because the variables may be contemporaneously related in that one series may lead to the other, or there may be a feedback relationship (Taio and Box, 1981). With time series data, this process is complicated as there is the need to model the interdependence between the series, as well as the serial dependence within component series (Chatfield, 2000).

The nature of multivariate models suggests that they should provide a greater understanding of the structures of a system as is utilises more parameters. Unfortunately, despite increasing computing power, this is often not the case for a number of reasons (Chatfield, 2000):

1. With more parameters to estimate there are greater opportunities for sampling variation to increase parameter uncertainty and affect forecasts.
2. With more variables to measure there is a greater opportunities for errors and outliers to appear.
3. The forecast of a dependent variable may require the forecast of explanatory variables which are not available, as such the forecast depends on the accuracy of the initial forecast.
4. Multivariate forecasts depend on having a good multivariate model, which cannot be guaranteed, as the model may be incorrectly identified, or may change over the period of fit.

Thus multivariate modelling is much more difficult to compute than univariate, and is more likely to be misspecified (Chatfield, 2000). As a result of these complications this model is not further explored in this thesis.

4.6. Forecasting

The ultimate aim of creating successful time series models is to be able to make forecasts as "the ability of the model to forecast future data values that have not been used in developing the model is more important and relevant than what has happened in the past" (Makridakis, 1978). These forecasting methods are primarily taken by predicting one step ahead of the calculated model. There are a number of basic forecasting methods which are described in Chapter 2, and more complex utilising the various ARMA models as described in Chapter 4.4.2. The simplest method assumes a linear trend and applies various smoothing methods.

4.6.1. Smoothing Methods

Smoothing methods originated in the 1950s and 1960s with the work of Brown in 1959, Holt in 1957 and Winters in 1960, and are seen as the extrapolation of univariate time series (De Gooijer and Hyndman, 2006). Pegels (1969) developed classification methods for displaying data with no trend, additive or multiplicative trends and trends with a seasonal effect in the exponential form. Gardner and McKenzie (1985) then extended this to include damped trends. Gardner and McKenzie (1988) developed a robust method for identifying exponential smoothing methods as it is common for the same model to be applied to all time series. The more common smoothing methods are outlined below.

Brown's (or Simple) exponential smoothing method is the simplest weighted average smoothing model. Introduced by Brown (1959) who states that "the estimate of the new average is updated periodically as the weighted sum of the demand in the period since the last review and the old average" (Brown, 1959:25). It has no allowance for trends or seasonality (Chatfield et al., 2001), and, therefore, is utilised for models which are essentially stationary such that $X_t=\beta_0+\epsilon_t$, β_0 is estimated at time t by

$$E_t=\alpha+(1-\alpha)E_{t-1} \tag{4.26}$$

where $0<\alpha<1$ is the smoothing constant of the exponentially smoothed values (Brown, 1959). The forecast of $\hat{X}_t(k)$ is given by

$$\hat{X}_t(k)=\alpha X_t+(1-\alpha)E_t \tag{4.27}$$

with the error of the forecast one step ahead given as $e_t=X_t-\hat{X}_{t-1}(1)$. The sum of squares (SSE), $\sum e_t^2$, of these errors is calculated and compared for different values of the smoothing constant α to see which yields the smallest sum of square errors (SSE) (Chatfield et al., 2001). This value of α is the best to use.

When a trend is present the simple model needs to be generalised. This is done through Brown's Double Exponential smoothing or Holt Winters Linear Exponential Smoothing method. These models take the form of $X_t=\beta_0+\beta_1 t+\epsilon_t$. In both cases a smoothing constant is calculated, with α being calculated for Brown's with smoothed values as $S_t=\alpha X_t+(1-\alpha)S_{t-1}$. The parameters α and β are calculated through Holt Winter's with the smoothing equations of

$E_t = \alpha X_t + (1-\alpha)(E_{t-1} + T_{t-1})$ and $T_t = \beta(E_t - E_{t-1}) + (1-\beta)T_{t-1}$ (Gardner and McKenzie, 1985). The SSE is calculated, and compared for different values of the constants, with the lowest SSE yielding the most suitable smoothing constants. However, as forecast models, linear trend models tend to overshoot the data in long term horizons (Gardner and McKenzie, 1985; Taylor, 2003), whereas a damped trend is more conservative and damps the trend according to the level of noise (Gardner, 1990). In this event, ϕ is less than 1, and when $\phi > 1$ an exponential trend is present (Harvey, 1994).

Should a linear trend with seasonal variation exist where the variance of the observations is non constant and increasing over time, a Holt Winter's Linear Multiplicative Seasonal Smoothing Method can be used. The model takes on the form $X_t = (\beta_0 + \beta_1 t) S_t + \epsilon_t$ where S is the seasonal index. For the trend $\beta_0 + \beta_1 t$ at time t and smoothing constant α (Winters, 1960),

$$E_t = \alpha\left(\frac{X_t}{S_{t-p}}\right) + (1-\alpha)(E_{t-1} + T_{t-1}). \tag{4.28}$$

For the gradient β_1 at time t with smoothing constant β :

$$T_t = \beta(E_t - E_{t-1}) + (1-\beta)T_{t-1}. \tag{4.29}$$

For the seasonal index at time t with smoothing constant γ :

$$S_t = \gamma\left(\frac{X_t}{E_t}\right) + (1-\gamma)S_{t-p}. \tag{4.30}$$

Thus the forecast of X_{t+k} at time t is:

$$\hat{X}_t(k) = (E_t + kT_t) S_{t+k-p}. \tag{4.31}$$

The SSE is calculated for different smoothing values of α , β and γ , and those which yield the lowest SSE are the parameters used for smoothing and prediction (Chatfield and Yar, 1988; Chatfield, 1989). Extensive work has been carried out with this method in production planning due to its robustness, but often including a damping trend of ϕ , such that t_{t-1} becomes ϕt_{t-1} (Miller and Liberatore, 1993; Taylor, 2003).

The Holt Winters Linear Additive Seasonal Smoothing model is applicable if the variance of the observations remains constant over time. It assumes a model of $X_t = \beta_0 + \beta_1 t + S_t + \epsilon_t$, where a linear regression model is fitted to the first observations:

$$X_t = \beta_0 + \beta_1 t + S_1 X_1 + S_2 X_2 + \ldots + S_p X_p + \epsilon_t \tag{4.32}$$

where the $X_i's$ are dummy variables. The SSE is calculated for different values of α, β and γ, and those which yield the lowest SSE are the parameters used for smoothing and prediction. The Holt Winters models are regarded as being robust, and an easy to use projection procedure, with the additional benefit of being automatic which is non time consuming (Chatfield and Yar, 1988). The Hold Winters model has been further extended by Gardner and McKenzie (1989) to include damped trends to give additional control over forecasting, and are "useful when the context of the situation suggests that any trends are likely to persist indefinitely" (Chatfield et al., 2001:156).

It is important to determine how good the smoothing and predicting methods are. Theil (1966 in Bliemel, 1973) suggested an inequality coefficient to measure this effectiveness. Assuming that the best prediction of the next observation is simply that it is the same as the last observation, i.e. $\hat{X}_{t-1}(1) = X_{t-1}$. If a set of predictions $\hat{X}_t$ (smoothed values) is available, it can be determined that it is better or non worse than a no-change model by calculating Theil's inequality coefficient (U) (Bliemel, 1973) :

$$U = \frac{\sqrt{\sum_{t=1}^{T} e_t^2}}{\sqrt{\sum_{t=2}^{T} (X_t - X_{t-1})^2}} \tag{4.33}$$

where *U=0* indicated that the model forecasts perfectly, *U<1*, that the model performs better than the no-change model, *U=1*, that the model performs as well as a no change model, and *U>1* that the model performs worse than a no-change model (Bliemel, 1973; Leuthold, 1975). From this, clearly it is desirable to have *U<1*, suggesting that the model produces reasonable forecasts.

4.6.2. Box-Jenkins Method

"The main purpose of fitting AMRA schemes is to project series forward beyond the sample period" (Johnston and DiNardo, 1997:231). In these forecasts there are two sources of error, error due to ignorance of future innovations, and error due to differences between true and estimated parameter values (Johnston and DiNardo, 1997). The Box Jenkins approach is applied as a methodological approach to apply to forecasting as described in figure 4.1 below.

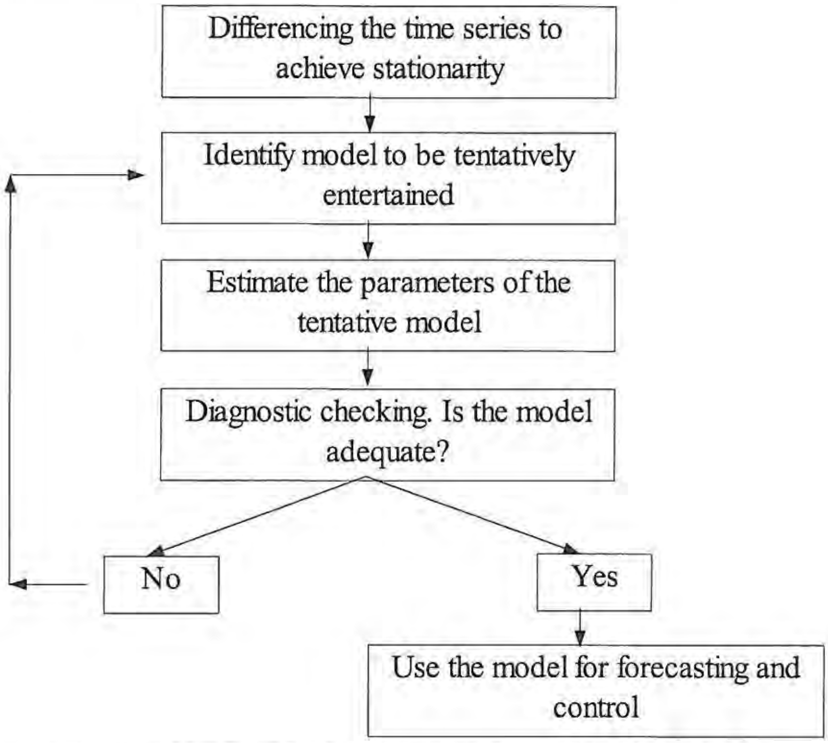


Figure 4.1: Box Jenkins Methodology (Maddala, 2003)

To make forecasts, it is assumed that the forecaster has determined the correct model and has the past and current values for e_t and the series y_t. If an AR(1) model is assumed of $y_t = a_0 + a_1 y_{t-1} + \epsilon_t$ and the first period is updated, we obtain

$$y_{t+1} = a_0 + a_1 y_t + \epsilon_{t+1}. \quad (4.34)$$

If the coefficients a_0 and a_1 are known, y_{t+1} can be forecasted based on

$$E_t y_{t+1} = a_0 + a_1 y_t \quad (4.35)$$

where $E_t y_{t+j}$ is the conditional expectation of y_{t+j} given the information available at *t*. In the same way the forecast at y_{t+2} conditioned on information at t is

$$E_t y_{t+2} = a_0 + a_1 E_t y_{t+1}. \quad (4.36)$$

Thus applying (4.31), so that the forecast of y_{t+1} is used to forecast y_{t+2}

$$E_t y_{t+2} = a_0 + a_1 (a_0 + a_1 y_t). \quad (4.37)$$

As such forecasts can be determined by an iterative process when the parameters are known. A forecast function can be determined as

$$E_t y_{t+j} = a_0(1 + a_1 + a_1^2 + \ldots + a_1^{j-1}) + a_1^j y_t \tag{4.38}$$

where j steps can be forecasted. The quality of the forecast declines as it is made further into the future. If we consider $\lim_{j\to\infty} E_t y_{t+j}$, we see that $\lim_{j\to\infty} E_t y_{t+j} = a_0/(1-a_1)$, so that for any stationary ARMA process, the conditional forecast of y_{t+j} converges to the unconditional mean as j converges to infinity (Enders, 2004).

Determining the error is important to outline the reliability of the forecast. The error of a forecast one step ahead of time is $e_t(1) = y_{t+1} - E_t y_{t+1} = \epsilon_{t+1}$, with the error after the jth step ahead forecast

$$e_t(j) = \epsilon_{t+j} + a_1 \epsilon_{t+j-1} + a_1^2 \epsilon_{t+j-2} + \ldots + a_1^{j-1} \epsilon_{t+1} \tag{4.39}$$

The mean of (4.35) is zero, as the expected value of $E_t \epsilon_{t+j} = 0$, thus the forecasts are unbiased estimates of each y_{t+j} (Enders, 2004).

Despite being unbiased, this does not necessary equate to accuracy. Considering the variance of the forecasts, if it assumed that the elements of ϵ_t are independent with variance of σ^2. Thus the variance of the error forecast is

$$Var[e_t(j)] = \sigma^2(1 + a_1^2 + a_1^4 + \ldots + a_1^{2(j-1)}). \tag{4.40}$$

Therefore the variance is an increasing function of *j*. Therefore greater confidence can be displayed in short term forecasts than long term. This method can be extended to higher order models, or any ARMA(p,q) (Enders, 2004). However, in practice the order of the ARMA process or actual values of the coefficients are not known, and as such they are estimated often through trial and error (Chatfield, 2000).

4.6.3. Forecast Evaluation

The model which fits the best does not necessarily make a good forecaster. This prompts the question of how to determine which model has the best forecasting performance. One method is not to use all the data in the model selection phase, and compare the different models on their ability to predict the unseen data. It is preferable that the forecast errors have a mean near zero and a small variance (Chatfield, 2000). There are a number of commonly used accuracy forecast measures, the most common being mean squared error, root mean squared measure, absolute error, and mean absolute percentage error (De Gooijer and Hyndman, 2006).

4.7. Conclusion

Time Series are time ordered observations, which are defined as either stationary or nonstationary. The objective of Time Series is to model the series as to enable predictions and forecasts to be made. This requires a stationary series, and as such methods have been developed to make nonstationary series stationary. These include differencing to remove trends, and transformations to remove second order nonstationarity, or fluctuating variances, with the objective of decreasing the autocorrelation to zero (Kendall and Ord, 1990; Chatfield, 2000).

Time Series models can be created as univariate or multivariate, but due to the complexities of multivariate techniques they are often negated (Chatfield, 2000). Univariate techniques are easier to forecast, using smoothing techniques, as well as estimating a model and the parameters. Autoregressive and moving average models are frequently developed, with ARMA models used as a combination. The parameters can be estimated using a least squares method to find the minimum error, which is then used in forecasting. In estimating a model and forecasting, a good model does not guarantee a good forecast, and as such, the forecasting ability should be tested against data unseen in the modelling stages.

CHAPTER 5
NEURAL NETWORKS

5.1. Introduction

There has been considerable interest in the development of Artificial Neural Networks for tackling a range of problems which may have a pattern recognition solution (Venugopal and Baets, 1994). Research has shown Neural Networks to have powerful pattern classification and recognition abilities (Zhang et al., 1998). Several distinguishing features of Artificial Neural Networks make them valuable and attractive for a forecasting task. First, as opposed to the traditional model-based methods, Artificial Neural Networks are data-driven self-adaptive methods in that there are few *a priori* assumptions about the models for problems under study (Tang and Fishwick, 1993). They learn from examples and capture subtle functional relationships among the data even if the underlying relationships are unknown or hard to describe (Zhang et al., 1998). Secondly, Artificial Neural Networks have the ability to generalise. After learning from data, they have the ability to predict from unseen data, even if it is noisy. Thirdly, Artificial Neural Networks are universal function estimators. It has been shown that Artificial Neural Networks can estimate any continuous function to any level of accuracy (Funahashi, 1989). The final advantage is that Artificial Neural Networks are non-linear as forecasting in the past has often been limited to predicting on a linear scale (Zhang et al., 1998).

The first Neural Network was developed in 1962 by Frank Rosenblatt (Beale and Jackson, 1998), and was called the Perceptron. It was modelled on the brain. As a learning machine, the Perceptron was a sensation in the 1960s, and began the era of Neural Networks. Using a learning rule, and a set of examples, it was able to produce target outputs for given signals. However, a major flaw emerged when it was unable to solve simple problems, such as an even or odd number of 'ones' in an input pattern or an "XOR" problem (Warner and Misra, 1996). As such, Neural Networks were neglected until 1985 with the discovery of new learning algorithms, and the construction of more powerful computers (Van Wezel and Baets, 1995; Nauck et al., 1997).

Neural Networks try to emulate the working of the brain, and this is where the founding research originated (Warner and Misra, 1996). The brain consists of a number of interconnected nerve cells called neurons, which influence each other by electrical signals. A signal is passed via its axion, and a signal is received over the connection between the axion and dendrite (Beale and Jackson, 1998). The connections between them are called synapses. Artificial Neural Networks use a similar principle, with non-linear computation elements called nodes as neurons, which receive inputs or information from other nodes or external stimuli. The information is processed with transfer functions (dendrites) which result in a final output (Venugopal and Baets, 1994; Zhang et al., 1998).

Different types of Neural Networks exist which can solve problems utilising a learning procedure which is dependent on the neural model and data, and require no underlying mathematical model (Nauck et al., 1997). Problems that can be solved include pattern recognition, pattern completion, detecting similarities between patterns or data and for interpolation or extrapolation (Nauck et al., 1997). As such Neural Networks have been growing in usage. A survey conducted by Wong et al., (2000) which reviewed journals from 1994 to 1998, revealed 309 business related applications for Neural Networks. This included industries such as the Accountancy, Banking, Human Resources Information Technology, Marketing and Production. Research is likely to continue in (Wong et al, 2000):

> production and finance [as they] usually involve a lot of difficult, complex, and non-linear applications, and neural networks technology is a tool that can handle these problems efficiently and effectively; the accessibility of raw data is relatively easy; and there are many potential real-world applications in the area of production/operations and can, therefore, simulate academics' and practitioners' interest in conducting research.

Neural Networks are particularly useful in demand forecasting, as they are non-linear, and involve many variables, and can ignore variables that do not contribute to the model (Church and Curram, 1996).

5.2. Design and Construction of a Neural Network

A Neural Network is constructed on the basis of a number of simple elements (Ton Su and Lin Hsieh, 1997). The most common Neural Network is the Mutli-Layer Perception (MLP) forward feedback (Zhang et al., 1998), and as such more attention is granted to it. They are used in forecasting due to their capability of input-output mapping. The MLP has many layers, or nodes. The first is the input layer, where external information is received, and the last is the output, where the problem solution is obtained. The input and output layers are separated by one or more hidden layers. Figure 5.1 gives a representation of a MLP with one hidden layer and one output.

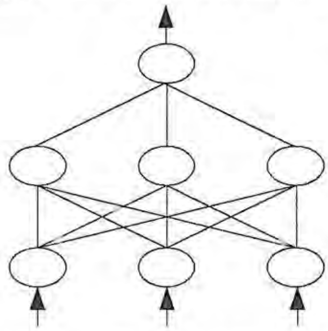


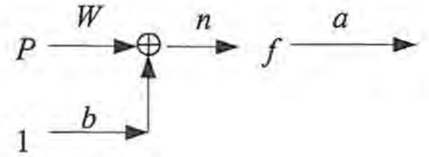
Figure 5.1: A Typical MLP Feedfoward Neural Network

To relate to a forecasting problem, the inputs are usually independent variables, and a relationship can be written as $y=f(x_1,x_2,...,x_p)$ where, $x_1,x_2,..,x_p$ are independent, and y is dependent. Extrapolating this to a time series scenario, $y_{t+1}=f(y_t,y_{t-1},...,y_{t-p})$ where y_t is the observation at time t. The benefit of Artificial Neural Networks is that it is easy to include predictor variables, and time lagged observations into one model (Zhang et al., 1998).

The most basic form of Artificial Neural Networks is the perception. It involves the process of pattern classification for linearly separable problems. This can be found by a hyperplane, in 2 or 3 dimensional space, of the form $W\,p+b=0$. A single node has inputs p of values $p_1,p_2,\ldots,p_n$ which are multiplied by the weights w of the neuron, $w_1,w_2,\ldots,w_n$, which correspond to the synaptic strengths of the neuron. The weighted input is summed such that $x_i \times w_i$, for $i=1,2,\ldots,n$ serves as the total combined input for the node. In addition, there can be a bias factor, b, which represents the intercept, similar to regression (Church and Curram, 1996), thus enabling the plane to be offset from the origin (Reed and Marks, 1999), which is also summed, such that the input n is obtained as in (5.1) (Patterson, 1990)

$$n=\sum_{i=1}^{r} w_i p_i+b = [w_1\cdots w_n]\begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix} = w \times p+b. \tag{5.1}$$

If the sum is large enough such that it exceeds a threshold value of T, the node fires into an activation function which produces an output a, such that $a=f(n)=f(w\,p+b)$. This is presented below diagrammatically (Patterson, 1990; Warner and Misra, 1996).



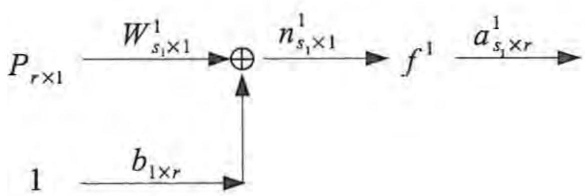
The system needs to be updated, and for this an update rule is used, such that the difference or error term e between the activation a and target t is 0. Thus the error term is denoted as $e=t-a$. If $e>0$, p is multiplied by e and added to the old weights, and e is added to b. If $e<0$, p is multiplied by e and subtracted from the old weights, and e is also subtracted from b. This results in an update learning rule, where the new weights are a function of the error term and the old weights and bias. This will continue for all $e\neq 0$ (Church and Curram, 1996; Burton, 2006).

In the MLF feedback, there are numerous layers, and the simple perception no longer works (Beale and Jackson, 1998), but the principle can still apply. Thus, the weight W^1_{ij}, is the strength of the connection between p_j and the neuron i. If the weight matrix is of size $s_1 \times r$, resulting in $n^1=W^1 p+b$ as the input to the first layer. Therefore $a^1=f^1(W^1 p+b^1)$ is the activation for the first layer. Thus if a^i is the activation from layer i, b^i is the bias in layer i. f_i is the transfer function in layer i, and W^i the weight matrix for the connections between layer $i-1$ and i, resulting in $a^i=f^i(W^i a^{i-1}+b^i)$. If there are m layers in the feed-forward Network,

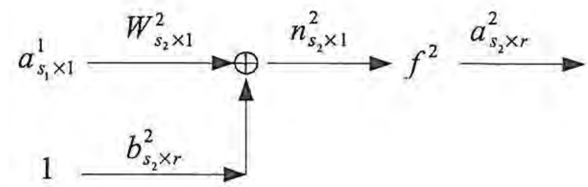
$a^1=f^1(W^1 p+b)$, $a^2=f^2(W^2 a^1+b^2)$, ..., $a^m=f^m(W^m a^{m-1}+b^m)$ (Burton, 2006).

The output from the first layer is passed onto the second layer, thus a Network receiving an $r-$ vector and sends it through M layers of neurons with transfer function f^i will have the following appearance (Burton, 2006).

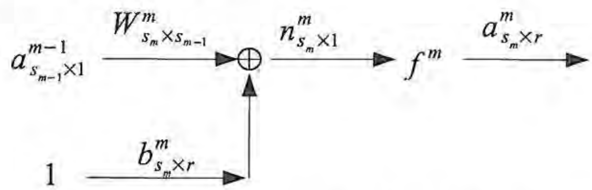
First Layer:



Second Layer:



which will continue until the m[th] layer is reached:



This is described as a $(r, s_1, s_2, \ldots, s_m)$ feed-forward Network. For example, the Network shown in figure 5.1 is a (3,3,1) feed-forward Network, where there are 3 input neurons, 3 neurons in the second layer, and 1 neuron in the output layer (Burton, 2006).

5.3. Network Architecture

Artificial Neural Network models are characterised by their properties, the structure of the Network (topology), how and when it computes and how the Neural Network learns to compute, based on the relationship between the input targets (Reed and Marks, 1999). Learning can be referred to as supervised or unsupervised. In supervised learning each output is paired with an input, and the trainer provides error information. In an unsupervised scenario, procedures capture regularities in the input values, without any additional information (Venugopal and Baets, 1994). This is useful, as unlabelled data are more readily available (Reed and Marks, 1999).

The core components of the Network architecture comprises of the number of layers, the activation function and the training algorithm. There are many approaches in determining the parameters of the Network architecture. These include the pruning method, polynomial time algorithm, canonical decomposition technique and Network information criterion (Zhang et al., 1998; Reed and Marks, 1999). The methods of finding the optimal architecture are complex, difficult to implement, and do not guarantee an optimal forecasting solution (Zhang et al., 1998), thus Neural Network design is often done through trial and error, based on experimentation (Warner and Misra, 1996; Darbellay and Slama, 2000). However, determining the best architecture is still crucial in the overall success of the Network (Hill et al., 1996).

5.3.1. Layers

The number of layers is problem specific, and it is important in the success of a Neural Network to capture the patterns in the data, and then to perform non-linear mapping between input and output variables (Zhang et al., 1998). Without the hidden layers, the Artificial Neural Network is equivalent to a linear statistical forecasting model. The literature is, however, in general agreement that more than 2 hidden layers produces no benefits (Reed and Marks, 1999). Although there is no consensus over one or two hidden layers, one layer is sufficient for most non-linear problems, but this may also require many hidden nodes, which makes it computationally heavy (Zhang et al., 1998). However, Elsken (1996) proved that adding additional layers results in a decrease in performance.

Determining the correct number of nodes, is crucial, as it influences the Network's capacity, learning ability, speed, and generalisation ability (Fujita, 1998). Too many nodes may result in over fitting, which is when the Network focuses on fitting the training data to the Network so well, that it learns of the random regularities contained in the training set (Tang and Fishwick, 1993; Schittenkopf et al., 1997), and as a result has a poor predictive ability on data which are outside the training sample (Van Wezel and Baets, 1995). However, if there are too few nodes, there may be insufficient power to model and learn the data. There are no theoretical bases for selecting the parameters, although a systematic method is advised. The most common method is via trial and error. Several rule of thumbs exist as well, such as at least 10 nodes, or $2n+1$ or $n/2$ as the maximum number of nodes to prevent over fitting where n is the length of the input data, but none of these work well over all problems (Zhang et al., 1998).

The number of input nodes corresponds to the number of input variables. There is also no suggested number of input nodes. However, sufficiently few is required which will reveal the underlying trends, and too few or too many may affect the learning or predicting capability of the Network (Fujita, 1998; Zhang et al., 1998; Hamid, 2004). The number of output nodes is simple to specify, it is simply the number of outputs, or results that are required.

The Network architecture is also characterised by the connections of nodes between layers. These connections fundamentally determine the behaviour of the model. For most applications the nodes in one layer are connected to all nodes in the next higher layer except for the output. It is possible to have more sparsely connected Networks, or have direct connections between the input and output which may be beneficial, as this may model the linear structures, which may improve forecasting ability (Zhang et al., 1998; Hamid, 2004).

5.3.2. Activation Function

The activation function is also known as the transfer function. It determines the relationships between inputs and outputs of the nodes and the Network. It normally aids in non-linearity which is advantageous, as this allows a continuous function (Beale and Jackson, 1998; Zhang et al., 1998). Any differentiable functions can be used as transfer functions, but in reality only a few are used due to their volatility around zero. The most common functions are (Zhang et al., 1998; Burton, 2006):

- Sigmoid (logistic) function

  $$f(x) = (1 + e^{(-x)})^{-1}$$

- hyperbolic tangent (tanh) function

  $$f(x) = \frac{(e^{(x)} - e^{(-x)})}{(e^{(x)} + e^{(-x)})}$$

- Sin or cosine function

  $$f(x) = \sin(x) \quad \text{or} \quad f(x) = \cos(x)$$

- Linear function

  $$f(x) = x$$

- Hardlim function

  $$f(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

with the logistics function being the most popular. The reason for this is that it acts like an automatic gain control, as it is a step function. Thus for small inputs, the slope changes rapidly, producing a large gain, and for large inputs the slope and thus gain is much less, therefore allowing the Network to accept large values while still remaining sensitive to small changes, effectively dulling the effect of outliers (Beale and Jackson, 1998; Hamid, 2004). Different transfer functions can also be applied at each layer, yet most Networks use the same function at each layer. There are some general rules for function use. Logistic functions have been suggested for forecasting problems, and most researchers use this. However, there is no rule for the transfer function on the output layer. Commonly the same function is used throughout. Important to note, is that a linear function cannot be used to forecast a time series which has a trend (Zhang et al., 1998). In addition, more than one linear layer is no more powerful than a single layer, as the linear function can be condensed into one operation. For example, if the Network is scaled by factor 2 in layer 1, and 5 in layer 2, it is equivalent to a once off scaling of 10 (Beale and Jackson, 1998).

5.3.3. Training Algorithm

The Neural Network training is a non-linear minimisation problem in which weights are adjusted to minimise the overall mean or total squared error between the actual output and desired output for all output nodes over all input patterns (Zhang et al., 1998). There are many choices for optimisation methods, and no method which guarantees a global minimum solution for all non-linear problems. As such all algorithms suffer from a local optima problem, so currently the best method is to select a method which gives the "best" optimisation result, given the true global solution cannot be found (Zhang et al., 1998). Numerous optimisation methods exist, with the most commonly used one being the back-propagation method, which is essentially the steepest gradient descent method (Roberts and Penny, 1997).

The back-propagation Network receives the input data pattern and passes it directly onto the hidden layer. Each element of the hidden layer calculates an activation value by summing up the weighted inputs and, then transforming the weighted input into an activity level by using a transfer function. Each element of the output layer is used to calculate an activation value by summing up the weighted inputs attributed to the hidden layer (Zhang et al., 1998). A transfer function is then used to calculate the Network output. Next, the actual Network output is compared with the target value. If a difference arises, i.e. an error term, the gradient-descent algorithm is used to adjust the connected weights. If no difference arises, the learning is complete (Ton Su and Lin Hsieh, 1997). The weights for a particular node should actually be adjusted in proportion to the error to the units in which it is connected, which is why back-propagating the errors through the net allows the weights between all the layers to be correctly adjusted (Beale and Jackson, 1998; Zhang et al., 1998). A simple process is proposed below (Ton Su and Lin Hsieh, 1997):

1. Calculate the output of each hidden node;
2. Calculate the output of each output node;
3. Calculate the error between target output and network output;
4. Calculate the modified gradient for output node;

5. Calculate the modified gradient for hidden node;
6. Modify the weight between output layer and hidden layer;
7. Modify the weight between hidden layer and input layer.

In the back-propagation method the learning rate should be specified, as it determines the size of the weight changes (Zhang et al., 1998). The steepest approach suffers from slow convergence, lack of efficiency and lack of robustness. Smaller learning rates slow the process, while larger rates may cause oscillation around the minima.

The back-propagation function is show below. If E_p is the error function for pattern p, t_{pj} represents the target output for pattern p on node j, while a_{pj} represents the actual output of the node. ω_{ij} is the weight of the nodes from node i to node j. The error is defined as

$$E_p=\frac{1}{2}\sum_j (t_{pj}-a_{pj})^2. \tag{5.2}$$

The activation for each unit j can be written as

$$n_{pj}=\sum_i \omega_{ij} a_{pi} \tag{5.3}$$

or the weighted sum as in the perception. The outcome from each unit j is the threshold function f_j acting on the weighted sum. This function f_j would be one of the activation functions mentioned above such that

$$a_{pj}=f_i(n_{pj}). \tag{5.4}$$

We can find the derivative of the error such that

$$\frac{\partial E_p}{\partial \omega_{ij}} = \frac{\partial E_p}{\partial n_{pj}} \frac{\partial n_{pj}}{\partial \omega_{ij}}. \tag{5.5}$$

Using the chain rule on the second term of (5.4)

$$\begin{aligned}\frac{\partial n_{pj}}{\partial \omega_{ij}} &= \frac{\partial}{\partial \omega_{ij}} \sum_k \omega_{kj} a_{pk} \\ &= \sum_k \frac{\partial \omega_{kj}}{\partial \omega_{ij}} a_{pk} \\ &= a_{pi}\end{aligned} \tag{5.6}$$

since

$$\frac{\delta \omega_{kj}}{\delta \omega_{ij}} = 0 \quad if\, i \neq j. \tag{5.7}$$

The change in error can be defined as a function of the change in net inputs

$$-\frac{\partial E_p}{\partial n_{pj}} = \delta_{pj} \tag{5.8}$$

and so (5.4) becomes,

$$-\frac{\partial E_p}{\partial w_{jk}} = \delta_{pj} a_{pj}. \tag{5.9}$$

Decreasing the value of E_p thus means making the weight changes proportional to $\delta_{pj} a_{pi}$ i.e.

$$\Delta_p \omega_{ij} = \eta\, \delta_{pj} a_{pi}. \tag{5.10}$$

Knowing δ_{pj} will enable us to know what to decrease E by. Using (5.7) and the chain rule,

$$\delta_{pj} = -\frac{\partial E_p}{\partial n_{pj}} = -\frac{\partial E_p}{\delta a_{pj}} \frac{\partial a_{pj}}{\partial n_{pj}} \tag{5.11}$$

Consider the second term and (5.3) again, then

$$\frac{\partial a_{pj}}{\partial n_{pj}} = f'_j(n_{pj}). \tag{5.12}$$

Now consider the first term in (5.10), and then differentiate E_p with respect to a_{pj}, such that

$$\frac{\partial E_p}{\partial a_{pj}} = -(t_{pj} - a_{pj}) \tag{5.13}$$

thus

$$\delta_{pj} = f'_j(n_{pj})(t_{pj} - a_{pj}). \tag{5.14}$$

This is useful, as we know both the target and the outputs, but the hidden layer is unknown as the targets are unknown. Thus if j is not an output, via the chain rule

$$\begin{aligned}\frac{\partial E_p}{\partial a_{pj}} &= \sum_k \frac{\partial E_p}{\partial n_{pk} a_{pj}} \frac{\partial n_{pk}}{\partial a_{pj}} \\ &= \sum_k \frac{\partial E_p}{\partial n_{pk} a_{pj}} \frac{\partial}{\partial a_{pj}} \sum_i w_{ik} a_{pi} \\ &= -\sum_k \delta_{pk} w_{jk}\end{aligned} \tag{5.15}$$

If the factors of (5.11) are substituted into (5.14), and applying (5.15),

$$\delta_{pj} = f'_j(n_{pj}) \sum_k \delta_{pk} w_{jk} \tag{5.16}$$

is obtained which represents the change in the error function with respect to the weights in the Network. The function is proportional to the errors δ_{pk} in subsequent units, so the error has to be calculated in the output units first, and then passed backwards through the algorithm to earlier weights to allow them to alter the connection weights. It is this passing backwards of the weights why it is called 'back-propagation (Nam and Schaefer, 1995; Beale and Jackson, 1998; Reed and Marks, 1999; Burton, 2006). It has been argued that back-propagation is computationally inefficient, thus making it time consuming (Sima, 1996), but with modern computing power, this is not the problem it once was.

In summary, inputs and targets are presented to the Artificial Neural Network as pairs, and the sum of squared errors over the whole training set is computed. If the sum of squared error exceeds the specified error target, the Artificial Neural Network adjusts the connection weights via back-propagation. Each pass through the Network is called a training epoch. The Artificial Neural Network then begins another training epoch until either the maximum number of training epochs is reached or the sum of squared errors reaches the specified error goal (Al-Saba and El-Amin, 1999). The training is said to be complete when either of this happens (Agrawal and Schorling, 1996). The maximum epoch is set to prevent infinite Network training in the event that a global minimum cannot be obtained.

5.3.4. Data Normalisation

Non-linear functions often limit the possible outputs of a node, for example, *logistic*, (-1,1), or *hardlim* (0,1). The data are normalised before the training process begins so that the data are in the range of the activation function, or roughly uniformly distributed (Hamid, 2004). This is done as it prevents the neurons from reaching saturation, as when this occurs there is little of no change in the output (Al-Saba and El-Amin, 1999). Even if a non-linear function is used, it is still useful to normalise the data to avoid computational problems, algorithm requirements, or to facilitate Network learning. There are four forms of normalisation (Zhang et al., 1998):

1. Along channel normalisation: Each input is normalise individually
2. Across channel normalisation: Elements are normalised across a data pattern
3. Mixed channel normalisation: Elements are normalised with a combination of along and

cross channel normalisation

4. External normalisation: All the training data are normalised into a specific range

The choice of method depends on the input vector. For a time series problem, an external normalisation is often the only method. For causal forecasting, across channel normalisation could be used, as the inputs are independent. Data normalisation has been found to be beneficial in terms of classification rate and mean square error, but this advantage diminishes as the Network and sample size increases (Shanker et al., 1996). Important to note is that the target values are also normalised, and as a result the output will be in a normalised form, so this must be rescaled to the original size. Levels of accuracy and other performance measures should be measured on rescaled data and not normalised data (Zhang et al., 1998).

5.3.5. Training Sample and Test Sample

A training set and test set are typically required for building a Neural Network forecaster. The training set is used for model development, and the test set is used to evaluate its ability at forecasting. Sometimes a third set, a validation set can be used to prevent the over fitting problem, or to determine the stopping point of the training process (Roberts and Penny, 1997). However, data are often valuable, and to split it into three sets, results in less data to be used for training which may make the Neural Network less robust (Zhang et al., 1998).

The division of the data into training and test sets is based upon the characteristics of the data, the data type, and the size of the data. The training and test set should be representative of the entire population of the data. Literature does not prescribe a specific training to test set ratio, but commonly, 90% vs 10%, 80% vs 20% and 70% vs 30% are used (Zhang et al., 1998).

Another important factor is the sample size. The amount of data is dependent on the Network structure, the training method, the complexity of the problem and the level of noise. The larger the size, the more accurate the results will be. Nam and Shaeffer (1995) found that as the sample size increases, Artificial Neural Network forecasting improves. In addition, the more complex the relationship, and the more noise, more data are required. The availability of data can be a limiting factor.

5.3.6. Performance Measures

There can be many performance measures, such as training time or modelling time, but the most important is the prediction accuracy that is obtained beyond the test set. There is also no generally accepted measure of accuracy. Accuracy is often defined as the difference between the actual result, and the predicted results. There are a number of different performance measures which have different advantages and disadvantages. The most frequently used are (Zhang et al., 1998; Reed and Marks, 1999):

- Mean absolute deviation (MAD) = $\sum |e_t| / n$
- Sum of squared error (SSE) = $\sum e_t^2$
- Mean square error (MSE) = $\sum (e_t)^2 / n$
- Mean absolute percentage error = $\frac{1}{n}\sum |\frac{e_t}{y_t}|(100)$

where e_t is the individual error, y_t is the actual value, and n is the number of error terms. The above mentioned error values are all absolute values, and cannot be used to compare different time series. Another problem with the MSE, is that it ignores the number of parameters that have to be estimated. Statistically, as the number of parameters to be estimated increases, the degrees of freedom decreases, thus increasing the chance of over fitting the model (Zhang et al., 1998).

5.4. Conclusion

Neural Networks have received a great deal of attention over recent years due to their ability in principle of producing good estimations of measurable data (Hill et al., 1996). They are modelled on the human brain, using a system of weights and biases utilised in conjunction with the back-propagation algorithm to produce a series of outputs. The algorithm is designed to minimise the errors between the predicted targets and the actual targets by adjusting the weights and biases according to the error (Beale and Jackson, 1998; Zhang et al., 1998).

This ability of Neural Networks being self learning to approximate functions (Hill et al., 1996), and requiring no initial function equation has made it ideal for forecasting complex non-linear relationships. As such it is increasingly being utilised in many areas, from medicine to commerce. This thesis exploits these characteristics by developing a Neural Network based on non-linear deodorant sales data to predict future consumer demand.

CHAPTER 6

METHODOLOGY

6.1. Introduction

This study attempts to create a consistent accurate demand forecasting system for use in the deodorant category of fast moving consumer goods by South African producers. This chapter presents and discusses the research methodology used in the study. Previously, Chapter 2 outlines the basic demand patterns that industries experience, with the subsequent two chapters detailing two techniques for predicting the demand; Time Series Analysis and Artificial Neural Networks. The literature notes that both methods discussed in this chapter are useful in different scenarios, but does not categorically state that either Time Series or Artificial Neural Networks is more effective in all situations. This chapter will highlight the methodology used in both of these methods and will show how these have been tested to make demand predictors with the data given.

6.2. Research Methodology

It is human nature to try and explain what we observe around us (Black, 1999). This exploration of the world is done through research. Research methodology is a "structured set of guidelines or activities to assist in generating valid and reliable research results" (Mingers, 2001: 242). The researcher aims to generalise the findings of a study to assist decision makers to draw conclusions about particular problems (Terre Blanche and Durrheim, 1999).

This is done in either a qualitative or quantitative framework, or a combination. The major difference between the two frameworks is that qualitative research admits the value-laden nature of the study, and actively reports their biases, as well as the nature of the information gathered (Creswell, 1994). The quantitative researcher views reality as 'objective', 'out there' independent of the researcher, something that can be measured objectively (Creswell, 1994). The researcher must make decisions based on how the research questions can be best answered in an investigation, and this often consists of a compromise between the ideals of good research and the numerous practical constraints that are present in the real-life research setting (Terre Blanche and Durrheim, 1999). Qualitative and quantitative techniques aim to study selected methods in depth, with qualitative attempting to understand categories of information that emerge from data, and quantitative to make broad and generalisable comparisons (Terre Blanche and Durrheim, 1999). This research aims to generalise a method of demand forecasting, and as such is conducted in the quantitative paradigm.

"The quantitative research is characterised by literature being used as a basis for advancing research questions and hypotheses" (Creswell, 1994: 24). These questions or hypotheses are then tested or answered through controlled experiments or statistical analysis. This is conducted within the positivist paradigm which makes the assumption that knowledge is objective, universal, true, and cumulative (Hart, 2005). The positivist assumes that this knowledge can be defined in a "theoretically neutral observation language" (Hughes and Sharrock, 1997: 43), thus suggesting that the research is independent of the researcher and research instruments (Lee, 1991). This positivist view "has its origins a school of thought from the philosophy of science, known as logical positivism" (Lee, 1991: 343). Logical positivism advocates a reductive view, that satisfies the natural science model of scientific research, dealing with facts and observable data (Lee, 1991; Babbie and Mouton, 1998).

Statistics is a discipline that uses a set of mathematical techniques that allow the researcher to make claims about the nature of the world using forms of principled statistical arguments (Terre Blanche and Durrheim, 1999). Using these techniques as a tool within the quantitative paradigm of research, the author aims to generate models that can predict demand and determine which predicts more accurately and consistently. It relies on data which are gathered, and from this patterns are looked for to generate models.

6.3. Research Population and Sampling Methods

The data for this study were obtained from Johnson and Johnson, South Africa. Johnson and Johnson is a manufacturer of health care products as well as a provider of related health care services for the consumer, pharmaceutical and professional markets (Johnson and Johnson, 2006). They provided data on the deodorant category in all the Pick 'n Pay Hypermarkets in the Western Cape, South Africa[8]. These data were captured by A. C. Nielson, a retail measurement organisation. A. C. Nielson utilises a (ACNielson, 2006):

8 South Africa's largest retail stores, which target customers on price and range (Pick 'n Pay, 2006).

Retail Measurement Services (RMS) [which] provide[s] continuous tracking of product sales to consumers, based on information gathered at the retail point-of-sale. Clients receive detailed information on sales, market shares, distribution, pricing and merchandising and promotional activities.

It is important to note that in this context the consumer is the retailer, and the individuals who purchase the products for individual consumption are known as shoppers. The data obtained by A. C. Nielson were obtained through point of sale bar code scans, thus can be considered accurate with a lesser chance of human error.

The population for this study would be regarded as all product sales in all stores countrywide. To collect accurate data for this population is not only impractical but is impossible. The impossibility arises because an informal economy exists in South Africa which is not monitored, and thus no data are available. The informal trade sector produces approximately 47% of the 1.9 million informal sector jobs in South Africa[9] (Devey et al., 2006) which provides an indication of the scale at which data are unavailable. In addition, all available data would present a research scope which is beyond that of this thesis. Thus, the author will use a more limited data set; that of the deodorants sold in the Pick 'n Pay Hypermarkets in the Western Cape of South Africa.

6.4. Main Study

6.4.1. Data

The data obtained from Johnson and Johnson were very detailed and were provided in the form of a Microsoft Excel Pivot table and have been assumed to be accurate. The data consist of monthly data over four years and included

- Promotion spent by brand and product
- Total sales by brand and product
- The number of products in the category in relation to the number of direct competitors.

The author began the research by separating the data into male and female deodorant brands, to remove any gender variables that may exist. The data were then manipulated to separate the variables into the factors which influence demand as described in Chapter 1. A section of the data is

9 StatSA is currently conducting a household survey to ascertain the true size of this informal economy (Denton, 2005).

shown in table 6.1below, with the descriptions as follows:

- Promotion: Total promotion expenditure in Rands per month
- Average Price: Average Price in Rands of the specific product per month
- Store Count: The number of Pick 'n Pay Hypermarkets the product is sold in
- Value: The total revenue in Rands generated by the product over the month
- Products: The number of products within each specific brand
- Product Total: The total number of products being sold in the stores at the same time, i.e. competitors
- Brand Total: The total number of brands been sold in the stores at the same time, i.e. Brand competitors
- Industry Price: Average Price in Rands of all deodorant product lines per month for all competitors
- Total Market: The total of all deodorants sold in the Hypermarkets over the month, i.e. It is the size of the industry
- Total Marketing: The total marketing expenditure in Rands spent by all players in the deodorant category
- Units: The total number of deodorants sold for the specific brand or product

	2002						**2005**
	January	**February**	**March**	**April**	**May**	...	**December**
Promotion	0	0	1,067,418	1,107,119	0	...	0
Average Price	9.096102	8.965899	8.147192	9.584766	8.77213	...	9.067916
Store Count	29	29	29	29	29	...	36
Value	134422.2	143337.8	208494.8	157535.2	166696.8	...	252196.9
Products	10	10	10	10	10	...	10
Product Total	182	186	175	191	193	...	235
Brand Total	34	34	33	33	34	...	43
Industry Price	10.99526	10.67223	10.92366	10.98908	11.10194	...	11.77042
Total Market	89847	90901	112741	100142	90947	...	192836
Total Marketing	738647	835955	5079072	3113616	2549008	...	2072403
Units	14778	15987	25591	16436	19003	...	27812

Table 6.1: A sample of Data

To ensure the validity of the research and ensure that there was no incentive for misrepresentation of data, all data sets are coded, *M1* to *M15* for the male deodorant product lines, and *F1* to *F14* for the female product lines. The codes are crucial to avoid brand biases, and to protect anonymity of firms. In addition, other more detailed studies were carried out on specific products, and specific stores, which are coded as S_i where i is the number of studies carried out.

A number of such assumptions are made to ensure that the different brands are comparable. It is assumed that price is a representation of quality and brand, with more expensive products being of a superior quality or fragrance than cheaper products. A brand value of the different products cannot be numerically ascertained, so we assume that all products carry an equal brand recognition. Promotion expenditure shows a large variance between months and between products and it is assumed that the promotion expenditure is spent optimally for all firms, that R1 spent by each firm has an equal effect on the consumer buying decision. It is also assumed that any benefits products earn through shelf positioning or otherwise are all obtained through promotional expenditure. It is also assumed that products of equal price are considered as perfect substitutes all other things being equal.

The data are used in a Time Series forecaster utilising only the 'units sold' variable, and into the Neural Network system all variables as detailed in section 6.4.2.2.1 below. The forecasts are made over 6 periods to ensure that the forecast for each deodorant product is consistent, and then the forecasts are compared over all 36 individual brands to ensure that the forecast method is robust for different products. A simple forecast, called the Previous Month forecast, is also generated where it is assumed that the number of products sold in the following month is equal to those sold in the previous month. This is the most simple form of forecast, and can act as a benchmark comparison.

6.4.2. Analysis

The analysis of the data was conducted in three parts: Time Series Analysis and through the construction of an Artificial Neural Network. The Microsoft Excel Pivot Tables were used to extract and manipulate the data into the form that was required. Microsoft Excel and Solver was used to run the Time Series analysis, and Matlab was used to create the Neural Network systems. The third part performs as reference point where it is assumed that future sales are the same as the previous month.

6.4.2.1. Time Series Analysis

Univariate Time Series methods were used, and as such data preparation is relatively simple, requiring just the series of the required forecasting variable. As total units sold is the desired prediction variable, the array of past units sold is required as a vector. As the literature suggests it is necessary to check for stationarity. Correlograms are generated to check for autocorrelations rapidly descending to zero to suggest stationarity. A unit root test is carried out to test for the presence of stationarity.

Nonstationary trends are to be transformed to stationarity through both differencing and transformations. Using stationary data, Box Jenkins methods are used to create an ARMA model for use in forecasting. Forecasts were also generated using Holt Winters Linear Multiplicative model, using Microsoft Excel Solver to optimise the values of the parameters.

Holt Winters Forecasts are made for a Holt Winters Linear Exponential smoothing model, with the parameters optimised using Microsoft Solver for the first $48-p$ observations where p is the number of forecasts to be made. This is repeated for $48-p+1$ observations where $p-1$ is the forecast number. This is repeated for all values of p. The forecasts are made onto unseen data, and an error rate is calculated, as

$$error = \frac{predicted\ \ total - actual\, total}{actual\, total}. \tag{6.1}$$

Similarly forecasts are made using an ARMA process. The data are tested for stationarity using a unit root test and correlograms on the first $48-p$ observations. A transformation is applied through differencing to smooth the data set to make it stationary, and the stationary data are extrapolated to make p forecasts. The forecasts are made on unseen data and the error rate calculated as in (6.1). A percentage error is calculated as there is a large range in total units sold between product brands, and as such a whole value error would be misleading and bias towards brands which sold fewer products. In addition the error rates can easily be converted into a financial variable. The forecast is also tested to determine if it is significantly different from the actual sales value. A differencing test is carried out, where the difference is tested at the null hypothesis of $H_0 : \delta = 0$, where there is zero mean difference between results. This is tested using the test statistic

$$t = \frac{\bar{D} - \delta}{s_d / \sqrt{n}} \tag{6.2}$$

where $\bar{D}=\frac{1}{n}\sum_{j=1}^{n} D_j$, D_j is the difference between results and $s_d^2=\frac{1}{n-1}\sum_{j=1}^{n}(D_j-\bar{D})^2$ with a t-distribution of $n-1$ degrees of freedom and a rejection region is created at $t_{n-1}(\alpha/2)$, and holds under the normal assumption (Johnson and Wichern, 2002).

6.4.2.2. Neural Network, MATLAB Analysis

6.4.2.2.1. Data Preparation

A successful analysis requires that the data are prepared accurately so that there are no errors which may give false results. The data preparation for the Neural Network is detailed below, with a full script of the code given in Appendix A.

Units were used as the target variable t_i, and the rest of the data was used as the input variable, p_i. However, it was decided that "Value" and "Total Marketing" would be removed from the training data. "Value" is a function of "Avg Price" and "Units". In addition, "Value" cannot be obtained without "Units", thus it does not make sense to use a variable in predicting stage that is ultimately only known after the prediction. "Total Marketing" was removed as, although potentially useful, it is not a variable that can be feasibly obtained for a live real world projection so, in the interests of practicality, it was ignored. All the remaining data that are used in p_i are obtainable from within the organisation. Total Market is also an unknown variable, but this can be predicted through Time Series analysis as explained in Appendix E.

The data were provided in a monthly form, so it was converted into a *sin, cos* variable, which depicts the data as a position on a circle, thus a January value will always carry the same position, thus enabling the Neural Network to determine if there is any seasonal pattern.

A memory function was incorporated into the Neural Network, as research suggests that consumers have a certain level of advertising retention or a carry over effect of advertising (Tull, 1965; Keller, 1987; Braun, 1999). Although numerically values of retention could be ascertained from the literature, it was assumed. The assumption in the Neural Network was that the promotion in month 3 was a function of the previous 2 months of promotion. It was assumed that the total promotion value was:

$$promotion = 0.4 \times promotion\ month_{n-2} + 0.6 \times promotion\ month_{n-1} + promotion\ month_n$$

The data were normalised using external normalisation so that components of the variable were between -1 and 1. This is simply achieved by a function shown below (6.2), which utilised the maximum and minimum of each variable

$$p_i = 2\left(\frac{p - min_p}{max_p - min_p}\right) - 1 \qquad (6.3)$$

This is done easily in Matlab using the built in function `premnmx` which accepts the input matrix p, and the target matrix t. It returns normalised matrices of p_n and t_n. The `premnmx` function determines the minimum and maximum of each row of p and t, and scales the data accordingly. It stores the variables $min\,p, max\,p, min\,t, max\,t$ which are used in rescaling the data. After the Neural Network has been trained using the normalised data, it is rescaled back to its original scale using the function `postmnmx`.

To test that predictions were consistently accurate for the same product over a number of months m, the system requests an input, n which is the number of months that it should predict for. The demand planning function utilises the first $m-n$ observations as training data, and then the final n observations as testing data. This repeats for all values, *1* to n For example in table 6.1, where $m=48$, as in the data frame in this thesis, and letting $n=3$.

Prediction	Training data - observations	Predict month
1	1 - 45	46
2	1 - 46	47
3	1 - 47	48

Table 6.2 : Example of Inputs and Targets

For each new prediction, all previous data are utilised. In a dynamic environment, where more data are constantly obtained, future predictions would train using growing data sets.

The built in Neural Network toolbox in Matlab utilises a random generator to seed the initial weights and biases for the back-propagation algorithm. Convergence to an exact number should occur, but due to the large range of data, the system cannot produce a consistently precise result. It does however lie within a range, so an average is taken over the number of simulations, y for each prediction. Thus to predict month $n+1$, the Neural Network was run y times and an average was taken. y is a variable controllable by the user of the system, but a value of 100 simulations was chosen for this thesis.

6.4.2.2.2. Network Architecture

A forward feedback (`newff`) network was utilised, as it has been shown to be the most effective at forecasting sales demand (Wong et al., 2000). Matlab offers a number of training algorithms which are described in table 6.3 below.

Training Algorithm	**Description**
trainb	Batch training with weight and bias learning rules
trainbfg	BFGS quasi-Newton back-propagation
trainbr	Bayesian regularisation
trainc	Cyclical order incremental update
traincg	Powell-Beale conjugate gradient back-propagation
traincgf	Fletcher-Powell conjugate gradient back-propagation
traincgp	Polak-Ribiere conjugate gradient back-propagation
traingd	Gradient descent back-propagation
traingda	Gradient descent with adaptive lr back-propagation
traingdm	Gradient descent with momentum back-propagation
traingdx	Gradient descent with momentum and adaptive lr back-prop
trainlm	Levenberg-Marquardt back-propagation
trainoss	One step secant back-propagation
trainr	Random order incremental update
trainrp	Resilient back-propagation (Rprop)
trains	Sequential order incremental update
trainscg	Scaled conjugate gradient back-propagation

Table 6.3: Description of the training algorithms in the MATLAB Neural Network Toolbox

The scaled conjugate gradient back-propagation (`trainscg`) algorithm has been used in other sales forecasting research (Wong et al., 2000). It proved successful and so was used. This algorithm requires significant memory to compute large Hessian matrices, but in the modern computational environment, this was not a concern. Other algorithms such as the Gradient descent back-propagation (`traingd`) and Levenberg-Marquardt back-propagation (`trainlm`) were tested to see if a better result could be produced, but it was found that this was not the case.

It was decided to use three layers layers in the Neural Network architecture, as more than that offers no additional benefit, and less does not predict non-linear relationships very well (Elsken, 1996; Reed and Marks, 1999). The tranfer functions in the layers were `tansig`[10], functions in layers 1 and 2 and a `purelin`[11] function in the third layer. The `tansig` functions were chosen as they are very sensitive around 0, in the range -1,1, and as the data had been normalised to a -1,1 range, this was particularly appropriate. The `purelin` function in the last layer, allows the output vector to obtain an arbitrary size. There is no deterministic method of determining the functions, but rather through a process of trial and error (Warner and Misra, 1996; Darbellay and Slama, 2000).

The network is in the structure form $[s_1, s_2, s_3]$ or [17,4,1] where S_i denotes the number of neurons in layer i. The input layer had 8 nodes to receive the 8 dimensional input vector, p_i. The two hidden layers had 17 and 4 neurons, with the output layer having 1 neuron to produce a 1 dimensional activation vector, a_i, which is the same length as the target vector t_i. The values for the hidden layers were obtianed through trial and error, as there is no definitive way of obtaining the optimal layer structure (Zhang et al., 1998). The layers s_1, s_2 were determined by manually adjusting the values, until the best error term was obtained. Other layers of values of 16, 5 or 18, 5 produce very similar results.

The target error goal during training was set to 0.005. It was not set to *zero* as this would likely result in overfitting. It was found that the network converges very rapidly to the goal, and as such the maximum number of epochs was set at 500. Training for too long also results in overtraining, so the maximum epoch prevents this if the error goal is not met (Zhang et al., 1998).

It is important to have a method of measuring the efficiency of the Neural Network. In this study, the error term was measured by calculating the error between the activation variable a_i in relation to the target t_i. As the network was run y times, which results in y activations, which are the predicted units sold, an error value was determined as follows:

$$error = \frac{mean\,(predicted\ total) - actual\ target}{actual\ target} \tag{6.4}$$

10 A hyperbolic tangent function

11 A linear function

The reason for this is because it is simple to understand, and in a business environment, the degree by which the target was missed is most important, as this can easily be translated into a financial variable.

The predictions of the different methodologies are compared for 29 deodorant brands to determine whether the forecasts are consistently accurate. Any significant anomolies are examined individually to determine whether there is a model flaw, or an anomolie in the data set itself which cannot successfully be accounted for in a model. A difference test is also carried out in a similar nature to as in the Time Series analysis to determine whether the forecast is significantly different to the true sales value.

6.4.2.3. Previous Month

This is the most simple method of forecasting, where it is said that sales in the following month are the same as the previous month. It acts as a minimum reference point that other forecasting methods should be able to out perform. A difference test is also carried out to determine whether there is a significant difference between forecasted sales and actual sales.

6.5. Factor Analysis

A factor analysis is carried out to determine any underlying relationships in the data that could be used to aid in forecasting. Factor analysis is a means of describing the underlying covariance relationship between two or more variables in terms of a few underlying factors. Factor analysis operates on the assumption that variables can be grouped by their correlations. Johnson and Wichern (1998) state that:

> If all variables within a group are highly correlated among themselves, but have small correlations with variables in a different group, then it is possible that each group represents a different underlying construct or factor.

The factor analysis attempts to describe the components by assuming that a vector **X** of p components has a mean of μ and a covariance matrix $\sum$. The factor model states that **X** is linearly dependent on unobservable random factors, $F_1, F_2, \ldots, F_m$ called common factors, and p variations $\epsilon_1, \epsilon_2, \ldots, \epsilon_p$ called specific factors and described as

$$\underset{(p\times1)}{X} = \underset{(p\times m)}{L} \; \underset{(m\times1)}{F} + \underset{(p\times1)}{\epsilon}. \qquad (6.5)$$

The communality is calculated as the portion of variance of the i^{th} variable contributed to the m^{th} common factor. Factor loadings are not always easily interpretable and it is common practise to rotate them until a simpler structure is obtained. The most common is the varimax rotation where the coefficients are scaled by the square root of the commonalities.

In this thesis a factor analysis was conducted using 4 factors and a varimax rotation using the statistical package R. In addition, the variable "Units" sold was removed, as the object of the factor analysis is to determine what relationships there are among the variables. This can then be applied to determine if there is an affect on units sold and so can be a useful indicator of the effects of changing various components of the marketing mix on the individual products.

6.6. Conclusion

This chapter described the methodology used in this study. Numerical predictions are required and as such the justification of quantitative methods for this study are explained. Two forecasting methods, Time Series techniques and Artificial Neural Networks are discussed in detail from preparation of data to interpretation of the results. These results are to be compared in an effort to determine whether a specific method is more reliable and whether it could practically be used in industry.

CHAPTER 7

PRESENTATION AND DISCUSSION OF RESULTS

7.1. Introduction

The previous chapter detailed the research methodologies, highlighted research theory, and explained the data collection and analysis used in this thesis. Once the data were organised, basic analysis was conducted with descriptive statistics to understand the nature of the data being used. This chapter details the results obtained from the relevant forecasts and makes inferences based on the available information in an effort to obtain a greater understanding. It focuses on comprehension of what marketing mix activities affects consumer demand, and how it can be applied to predict product demand accurately.

7.2. Results

The results are first discussed focusing on each individual model, but an analysis is more effective when compared against each other as it demonstrates which forecasts are most successful. Displaying the results for projections would not yield clear results, so detailed explanations are given for select results. A full set of projection results are available in Appendix B, including Theil's U-values and differencing test results in Appendix C. Although 29 data sets were available only 12 male and 12 female data sets were considered sufficient to make comparable forecasts as the other data sets did not contain a full array of data over 48 months.

7.2.1. Time Series

Both Box Jenkins ARMA and smoothing techniques were investigated to make forecasts. Using the Box Jenkins ARMA technique, forecasts could not be made. The data were found to be nonstationary using a unit root test, but could not be made stationary through differencing or transformations. As a result this method was no longer considered as a suitable forecasting method. Using the Holt Winter Linear Exponential Smoothing method, a Theil's U-value of greater than 1 was obtained for all of the Time Series projections. This suggests that the models perform worse than a no change model. Although discouraging in terms of predictive accuracy, the results are still presented here.

A difference test was carried out on the Time Series forecasts to determine whether there was a significant difference between forecasted sales and actual sales. The rejection region for the male difference test is $t_{71}(0.05/2)=1.666$, and for the female difference test, $t_{69}(0.05/2)=1.667$. The test failed to reject the null hypothesis, at the 5% level of significance, 58% of the time for male and 50% of the time for female deodorant products, suggesting that the mean of the forecasts is not significantly different from the actual results only just over half of the time. It may also indicate that over a 6-month prediction period the total forecast is reasonably accurate and 'over predictions' are compensated by 'under predictions', thus ensuring that stocks do not build up or run short. However, when the predictions were tested as a group for all male forecasts and all female forecasts for a difference from zero, there was sufficient evidence to reject the null hypothesis, suggesting that the mean is significantly different from zero for both male and female products, with t-values of -2.6406 and -2.333, respectively. Thus, these t-tests indicate that the Time Series forecasts produce forecasts which are on average not significantly different from the actual sales for individual brands forecasts, but are significantly different when forecasting all brands together. In conjunction with Theil's U-statistic indicating that the forecasts are worse than a no change model, it can be concluded that Time Series forecasts used are not accurate.

7.2.2. Previous Month Forecast

The Previous Month forecast was also tested for a mean zero differences using the difference test. The test failed to reject the null hypothesis, at the 5% level of significance, 66% of the time for male and 41% of the time for female deodorant products. When tested as a group on all male and female data, the null hypothesis was rejected indicating that the mean of the forecasts is significantly different from the actual results. This suggests that the Previous Month forecasts are not consistently reliable, much like the Time Series forecasts.

7.2.3. Neural Networks

The Neural Networks simulations were run generating forecasts for months 42 to 48. 100 simulations were used for each product and an average was generated over the 100 simulations. This was done in an attempt to stabilise the large variance produced in the forecast which was generated as a result of the random variable. It was decided that 100 repetitions was sufficient to create a stable average. Figure 7.1 displays the output of 100 forecasts for F9 at the 5th forecast month. The large variance is clear from the spiking of the prediction line, but the average prediction still produces an error of -3.3%. This error value is considered to be good as it is significantly lower than industry forecasting errors which have a range of 10% to 70% (Jeunet, 2006).

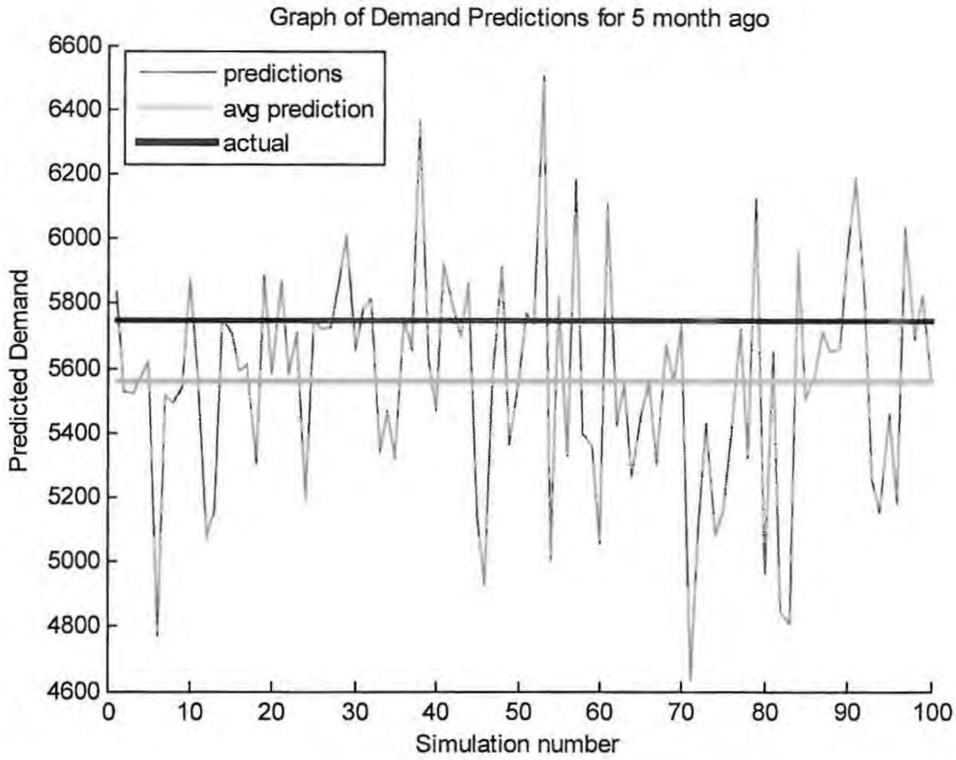


Figure 7.1: Matlab Output Forecast Example

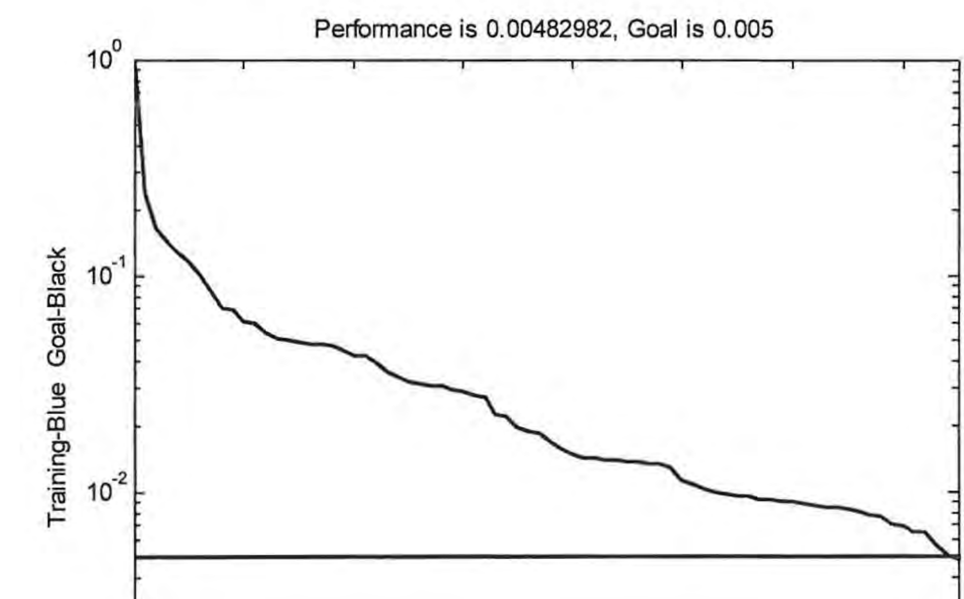


Figure 7.2: Matlab output of Epochs

Of the 21,600[12] simulations, approximately 65 epochs were required for the Neural Network to decrease the error to 0.005, an example is shown in figure 7.2. This is considered to be a rapid rate of decrease, suggesting that there is a definite relationship between the variables and units sold.

The removal of outliers in the forecasts was considered. However, the upper and lower limits of the forecast are approximately equal, thus nullifying the individual effects of the outliers. As such, it was decided to retain them in the analysis, as opposed to manipulating the results. Manipulating the results may have had a negative effect in other situations.

A mean difference test was carried out in an identical manner to the Time Series forecasts. The full results are shown in Appendix C. The test failed to reject the null hypothesis, at the 5% level of significance, 66% of the time for male deodorant products, and 58% of the time for female data, suggesting that the mean of the forecasts is not significantly different from the actual sales 66% for male deodorants and 58% for female deodorants. This suggests that the forecasts are accurate over a 6 month period for the individual brands. When the predictions were tested as a group for all male forecasts and all female forecasts for a difference from zero, there was insufficient evidence to reject the null hypothesis, suggesting that the mean is not significantly different from zero for both male and female products, with t-values of -0.042 and 0.486, respectively. Thus, these t-tests indicate that the Neural Network forecasts produce forecasts which are on average not significantly different from the actual sales for individual brands forecasts, as well as a grouped brand forecast. This suggests that the Neural Network forecasts produce more accurate forecasts than the Time Series forecaster based on a single variable, namely Units sold.

7.2.4. Joint Results

In determining whether different forecasting techniques are more effective, they are compared against each other using individual percentage errors and graphs to determine what techniques are preferable in individual cases, and then the mean square error is determined and compared over 36 cases to determine consistency. This comparison was made for the different forecasting techniques used on M1 over all 6 forecasts and given here as an example. A graph of the results of M1 are shown in figure 7.3.

---

12 36 products, 6 month forecasts, and 100 simulations multiplied to produce 21,600 simulations

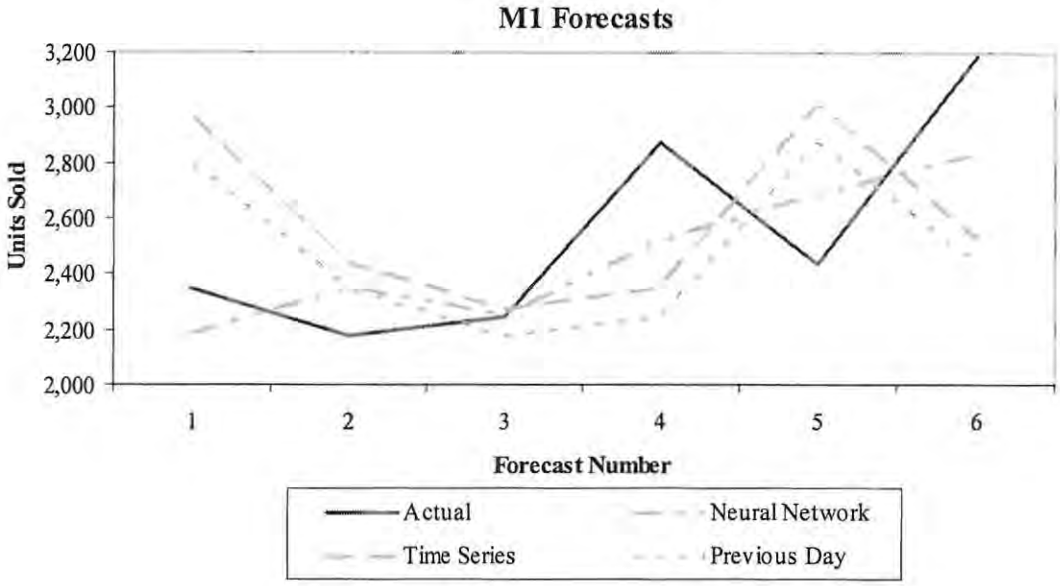


Figure 7.3: M1 Forecasts

Neural Network forecasts were found to produce the most accurate forecast 58% of the time, Time Series 16% of the time, and using the Previous Month forecasts were found to be more accurate than the other techniques 26% of the time for male products. For female products, 49% of the time the Neural Network produced the best forecast. This was followed by the Previous Month forecast at 31% of the time, and finally Time Series produced better forecasts 20% of the time. This is summarised in table 7.1 below

	Percentage of Time best Forecasting Method			
	Neural Network	Time Series	Previous Month	n
Male	58%	16%	26%	72
Female	49%	20%	31%	69[13]

Table 7.1: Best Forecasting Method Numerically

Although this creates an indication of what the preferred method of forecasting is individually, it does not provide an indication of what the error rate is to determine how successful the forecasts are when not optimal. For this the sum of squared errors (SSE) is calculated, such that $SSE = \sum e^2$. The sum of squared errors is calculated using the percentage error of the forecasts in an attempt to standardise the forecast errors. The variance between products sold is large and as such a small percentage error may equate to a large numerical error for certain products. The sum of squared errors for each forecast method is presented below in table 7.2.

13 69 units were used, as the final 3 forecasts for F5 were removed, as the product was removed from the stores in the final 3 months of forecast.

	Sum of Squared Errors (SSE)		
	Neural Network	Time Series	Previous Month
Male	13,680.62	36,994.21	23,715.02
Female	25,593.2	38,013.96	116,470.5

Table 7.2: Sum of Square Errors

It is interesting to note that although the Neural Network forecast did not always produce the best forecast, it appears to be consistently closer to the true value as it has the lowest SSE.

Industry has a target projection error of 20%, either above or below the true sales value (Jeunet, 2006). Table 7.3 shows the percentage of error greater than the industry target for the three forecasting techniques.

	Error greater than 20%		
	Neural Network	Time Series	Previous Month
Male	11%	35%	22%
Female	9%	25%	26%

Table 7.3: Percentage error above 20%

This suggests that Artificial Neural Networks predict consistently closer to the true value than Time Series and Previous Month forecasts. Upon closer examination of the error values, it is interesting to note where the anomalies occurred on the Neural Network forecasts.

Male Deodorant Forecasting Anomalies:

Product	Forecast Anomaly month	Explanation
M3	1	The Neural Network failed to determine that the price decrease of R1.50 would result in a sales increase of 11,000. It only predicted an increase of 4,000 units. When the data set changed substantially, the Neural Network failed to immediately register these large structural changes. However, it adjusted immediately for the following forecasts, and was still a substantially better estimate than both Time Series and Previous Month forecasts which predicted approximately equal sales values.
M12	3	The Neural Network over predicted the effect on demand of increased stores and decreased price. However, this was self corrected again in the following prediction.
M13	2 and 3	The Neural Network failed to take into account the effect of the industry average price decrease which had a negative effect on the promotion expenditure. As the product price decreased to reflect the industry price decrease, demand forecasts improved.
M14	4 and 6	The Neural Network failed to estimate accurately due to the sudden structural change in industry price. It recovered for the 4^{th} forecast.
M15	4 and 6	The Neural Network failed to determine the structural decrease and increase in forecasts 4 and 6, respectively, as price decreased by R1.20 for forecast 4 and increased by R1.30 between forecasts 5 and 6.

Female Deodorant Forecasting Anomalies:

Product	Forecast Anomaly month	Explanation
F4	5 and 6	The Neural Network was unable to determine that sales were deceasing, as a result of an industry average price decrease and forecasted a value similar to what had occurred the previous year under similar price and promotion conditions. This suggests that Neural Networks are poor predictors if there is a rapid structural change in either price or promotion.
F5	2 and 3	The values are so low as the product was in the process of being discontinued, and was removed entirely the following month.
F9	3 and 4	Again the Neural Network struggled when there was a rapid structural change, with a large decrease in industry average price, and an additional 15 market competitors, but at predictions 5 and 6 it had re-learnt the pattern and had adapted and the error decreased to acceptable levels of 14% again.

Given these results in tables 1.1, 1.2 and 1.3, it can be suggested that the Neural Network appears more effective. The Neural Network does struggle, however, when there are structural changes, such as large unexpected price decreases. Neural Networks are also trained within a range, and are accurate within the range of previously observed variables. When variables outside this range are presented the Neural Network extrapolates and loses accuracy but it does improve its accuracy over time. With human intervention, forecasts could be improved too. For example, if there was a large price decrease, the forecast could be artificially increased.

The Time Series and Previous Month models operate on the assumption that only past sales data are relevant, and as such large variations cannot be explained except as a random occurrence. This makes them both poor forecasting models as they do not take into account elements of the marketing mix such as price and promotion. They also fail to react to seasonal changes. Like the Artificial Neural Network, these methods fail to account for large structural changes as the parameters are not utilised in the model.

7.3. Factor Analysis

A factor analysis, as described in the methodology, was carried out to determine whether there are any relationships between the market variables which could be exploited to manipulate demand, and this knowledge could then be used to improve demand forecasts.

There are potentially 29 data sets to examine, but only M1, and M11 were chosen for male data to study as M1 has comparatively little promotion expenditure, and M11 has a significant amount of promotion expenditure. The difference in promotion expenditure was selected as it represented a major difference in the data sets and so would allow for a more representative study. Similarly F1 and F11 were studied as the female data samples. The results of the male and female factor analysis are shown in Appendix E and discussed below.

Male Deodorant Brand Factor Analysis

This section offers an interpretation of the results of the factor analysis shown in Appendix D.

M1: The model is significant at the 5% level of significance, so inferences can be made based on the results. Factor one is a weighted average of total products available in the market, and products in this brand, suggesting a competition variable. Factor 2 is a weighted average of product price, and industry average price, suggesting a pricing variable. Factor 3 is approximately a weighted average of market size and number of stores, suggesting a market size variable. Factor 4 is weighted on the promotion variable and the number of products and brands, suggesting a promotion – competition variable.

M11: The model is significant at the 5% level of significance, so the results do allow for inferences to be made. Factor 1 is comprised of a weighted average of store count and total market size for M11, this suggests that the first factor is a market size variable. Factor 2 is approximately a weighted average of product price and average market price, suggesting that the second factor is a pricing variable. Factor 3 is a weighted average total of the total brands and total products suggesting that it is a competition variable. Factor 4 is weighted on promotion expenditure suggesting that a promotion variable is present.

7.3.1. Female Deodorant Brand Factor Analysis

This section offers an interpretation of the results of the factor analysis shown in Appendix D.

F1: The model is significant at the 5% level of significance. Factor 1 is a weighted average of the number of stores, products in range and total number of brands, suggesting an availability variable. Factor 2 is heavily loaded on the average price, suggesting a price variable. Factor 3 is loaded on the number of products, suggesting that the number of products available is a variable. Factor 4 is a weighted average of the number of stores, and total market size with a contrast on promotions, suggesting that there is an inverse relationship between promotions and the numbers of stores the product is sold at.

F11: The model is significant at the 5% level of significance. Factor 1 is a weighted average of store count, brands, and products in line. This suggests that there is an availability variable. Factor 2 is heavily loaded on the average price suggesting a pricing variable. Factor 3 is a contrast between total sales and price, suggesting a trade off between products sold and price. Factor 4 is loaded on promotion suggesting a promotion variable.

7.3.2. Factor Analysis Conclusion

The male factor analysis based on these results suggests that the male product market is fairly similar and could comprise of a pricing variable, a competition variable and a promotion variable. Female deodorant brand data do not produce as constant and as clear groupings as does the male factor analysis. This can suggest that the buying pattern between the groups is different. F1 possibly suggests that females are likely to change stores due to promotions, but this is not fully supported by F11. The male deodorant brand factor analysis suggests that there is a pricing variable, but both female examples suggest a contrast between price and total sales, suggesting that woman may be more price sensitive than men. This factor analysis does not, however, conclusively prove any relationships between variables or differences between male and female, merely that the relationships may exist. However, the factor analysis of individual products can aid in determining how to manipulate the marketing mix as interactions are exposed which could be manipulated to yield greater sales.

7.4. Research Limitations

Although this research attempted to create comparable forecasts, there are factors that may affect the validity of the comparisons. These factors include data limitations and limitations of practical accurate forecasts.

The data set obtained was assumed to be accurate and may present a limitation in itself. The data obtained from ACNielson may be subject to human error in programming the systems that capture the data. In addition the data obtained for the promotional expenditure of firms on different brands cannot be verified, and firms may have misrepresented their promotional expenditure in an effort to retain corporate privacy.

One of the intentions of this research was to create a practical forecaster that could be applied in a real life environment. This reality of the research leads to limitations in this objective. The Neural Network forecast requires a large number of variables, and in reality many of these cannot be ascertained with certainty to make the forecasts. As a result these variables must be predicted as well, and forecasting with predicted values lowers the accuracy. The variables that need to be forecasted to use within the Neural Network include total market size, number of competitors, average price, and number of competing brands[14]. These can be predicted accurately as average price is almost constant, the number of competing products and brands adjusts incrementally and so is predictable. Total market size follows a seasonal trend, and can be predicted as a result. However, any forecast errors which are used in the Neural Network are likely to accentuate the error produced in the Neural Network. In this thesis known values were used in the Neural Network forecasts to ensure that only the forecasting ability of the Neural Network was being measured.

7.5. Future Research

This research has tested different forecasting techniques of predicting total sales within a shopping chain. Having found Artificial Neural Network methods to be the most successful at predicting future demand, future research can be carried out to determine the flexibility of Neural Network forecast models in more specific scenarios. Future research can be carried out to determine how much training the Neural Network requires to produce accurate forecasts. In addition it can be tested to determine whether Neural Network models are sufficient for predicting sales of specific products as opposed to brands, and whether sales can be predicted accurately in specific stores. This

14 Forecasts of these are shown in Appendix E

is briefly investigated below giving an indication of what future research may find.

This research has established that Neural Networks produces good forecasts but it should be tested to determine how many training sets are required before the Neural Network runs efficiently. All data sets forecasted utilised 4 years of data, i.e. 48 monthly points, so had a minimum Neural Network training time of 42 months. However, practically 4 years of data are too many to create an acceptable industry forecaster, so research should be conducted to determine what the minimum acceptable training length is. A forecaster was created with less than 10 training points, and appeared to train very rapidly to an acceptable error of less than 20%. This was tested on M5, and it forecasted to an error of 13% with only 6 training points, thus suggesting that Neural Networks forecasters can be trained very rapidly. Further research, therefore, should be carried out to determine how many months of information are necessary before Neural Network forecasts can be relied upon.

The research in this thesis has created a general forecast for an entire product range, for example, "Shield", and has ignored that it is sold in different variants of roll on, aerosol, different volumes, and different fragrances. To further test the ability of Neural Networks on a specific product, forecasts were calculated for S_1, a specific product variant including a specific fragrance and volume, with results as in table 7.4.

Forecast month	Actual	Neural Network Forecast	% Error
1	1627	1272	-21.82
2	770	923	19.87
3	856	860	0.46
4	1153	1236	7.2
5	1374	1224	-10.92
6	873	1269	45.36

Table 7.4: Specific Product Forecast for S_1

This result suggests that forecasts on individual product lines can be generated with forecasts within the target range of a 20% error. The anomaly at forecast number 6 is due to the product being discontinued, as the months preceding this it was sold at fewer stores, and totally discontinued 2 months later. This suggests that the sales were lower due to a lack of stock as the product was in the process of being discontinued. This result suggests that Neural Networks can be developed for all individual product lines to produce accurate forecasts for each individual product.

Neural Networks may not only be useful for the producers, but for individual stores as well. It is possible that stores can use a Neural Network forecasting technique to determine what stock of products they should purchase. A test was conducted for S_2, a specific product at a single Pick 'n Pay store, with the results shown in table 7.5.

Forecast month	Actual	Neural Network Forecast	% Error
1	63	247	292.06
2	47	496	955.32
3	52	236	353.85
4	68	167	145.59
5	63	118	87.3
6	19	111	484.21

Table 7.5: Forecast at a Specific Store S_2

The results are very poor, strongly suggesting that this Neural Network system is not sufficient for forecasting such small sales volumes. Possibly a more effective Neural Network could be developed using a different Neural Network architecture, but further research should be carried out to determine if improved specific product and store forecasters could be developed.

Future research could be carried out to determine whether Neural Network forecasters can predict sales of specific products, within specific stores, and of specific products within stores. This will enable both manufactures and retailers to forecast sales more accurately and reduce inventory costs. This research has been carried on the deodorant category, and further research should be carried out to determine whether these methods would be successful in other fast moving consumer good product lines as well.

7.6. Conclusion

This research explored the plausibility of developing a reliable forecasting method for the fast moving consumer goods industry. Three methods were utilised: a Time Series forecaster, a Previous Month forecaster, and an Artificial Neural Network forecaster.

Univariate Time Series and Previous Month forecasts used in this thesis assumed that there is a relationship between future sales and past sales. Neural Networks assumed that there is a relationship between a number of variables that influence sales. The Neural Network developed is designed to determine these relationships, and be able to forecast future sales based on the known relationships.

The Neural Network forecaster was found to be the most reliable, producing the forecasts which were most often not significantly different from the actual value. The Neural Network also produced forecasts which were most commonly closest to the true value, as well as the total smallest error. In addition it produced forecasts that were most often in the target range of a 20% error.

From this research it suggests that Neural Networks can be developed to produce consistent accurate demand forecasters. An awareness of the limitations of the Neural Network is necessary. It has been shown to fail to accurately forecast demand when presented with data which are outside of the training set. Practitioners should be aware that structurally different data may results in a poor forecast.

CHAPTER 8

CONCLUSION

This thesis has explored various methods of forecasting demand. The need for this research has arisen because firms are unable to accurately forecast consumer demand, leading to financial losses. This inability to forecast demand is problematic as it carries an expense in inventory management. Firms incur a cost in ensuring that inventory is available for the consumer. This thesis has examined the nature of demand forecasting, factors of marketing that affect it, methods of predicting demand and the results of those methods when carried out on a specific data set.

The literature discussed has highlighted a number of elements that can be considered when forecasting. It has been suggested that the marketing mix has the ability to influence consumer's purchasing behaviour, and that an understanding of this offers firms the ability to both influence demand, as well as a greater ability to predict demand based on this knowledge. This thesis discussed this literature, highlighting the probable relationship between different factors of the marketing mix and the changes in demand that result from changes in these factors.

The aim of this thesis was to examine models of demand forecasting in an attempt to find a successful method of a demand planning. Two methods were considered in determining an accurate forecasting model. Time Series techniques are used to develop forecasts based on previous sales data, and an Artificial Neural Network was created as a learning system to model relationships between variables. The Time Series technique did not prove as successful as using Artificial Neural Network as a forecaster.

Artificial Neural Network forecasts produced the most accurate forecasts over all data sets. The major benefit of the Neural Network is that it was able to use all sales influencing variables in the self learning model to accurately forecast demand.

Neural Networks have been shown to be a consistently accurate means of forecasting future sales. Although regarded as a "black box method" (Van Wezel and Baets, 1995) it can be effectively used in industry to improve sales forecasting methods enabling firms to reduce costs.

REFERENCES

ACNIELSON, 2006. **Our Products**. [On-line] http://www2.acnielsen.com/products/index.shtml (accessed 12 September 2006)

AGRAWAL, D. AND SCHORLING, C., 1996. "Market Share Forecasting: An Empirical Comparison of Artificial Neural Networks and Multinomial Logit Model". **Journal of Retailing**. 72(4): 383-407

AL-SABA, T. AND EL-AMIN, I., 1999. "Artificial Neural Networks as Applied to Long-Term Demand Forecasting". **Artificial Intelligence in Engineering**. 13(2): 189-197

BABBIE, E. AND MOUTON, J., 1998. **The Practice of Social Research**. Oxford University Press: Cape Town

BEALE, R. AND JACKSON, T., 1998. **Neural Computing: An Introduction**. Institute of Physics Publishing: Bristol

BENNETT, A.R., 1997. "The five V's - A Buyer's Perspective of the Marketing Mix". **Marketing Intelligence and Planning**. 15(3): 151-156

BERKOWITZ, E.N., KERIN, R.A., HARTLEY, S.W. AND RUDELIUS, W., 2000. **Marketing (6e)**. Irwin McGraw-Hill: New York

BLACK, T.R., 1999. **Doing Quantitative Research in the Social Sciences**. Sage Publications: London

BLIEMEL, F., 1973. "Theil's Forecast Accuracy Coefficient: A Clarification". **Journal of Marketing Research**. 10(4): 444-446

BOX, G.E.P. AND JENKINS, G.M., 1968. "Some Recent Advances in Forecasting and Control". **Applied Statistics**. 2(17): 91-109

BRAUN, K.A., 1999. "Postexperience Advertising Effects on Consumer Memory". **The Journal of Consumer Research**. 25(4): 319-334

BROWN, R.G., 1959. **Statistical Forecasting For Inventory Control**. McGraw Hill: London

BROWN, S., 2002. "Vote, Vote, Vote for Philip Kotler". **European Journal Of Marketing**. 3(3): 313-324

BURTON, M., 2006. **Neural Networks**. Unpublished: Grahamstown

CATARELLI, E.J., 1980. **Developing New Products and Repositioning Mature Brands**. John Wiley and Sons: New York

CHATFIELD, C. AND YAR, M., 1988. "Holt-Winters Forecasting: Some Practical Issues". **International Journal of Forecasting**. 37(2): 129-140

CHATFIELD, C., 1989. **The Analysis of Time Series: An Introduction**. Chapman and Hall: New York

CHATFIELD, C., 2000. **Time Series Forecasting**. Chapman and Hall/ CRC Press: Florida

CHATFIELD, C., KOEHLER, A.E., ORD, K. AND SNYDER, R.D., 2001. "A New Look at Models for Exponential Smoothing". **The Statistician.** 50(2): 147-159

CHURCH, K.B. AND CURRAM, S.P., 1996. "Forecasting Consumers Expenditure: A Comparison Between Econometric and Neural Network Models". **International Journal of Forecasting**. 12(2): 255-267

CONSTANTINIDES, E., 2002. "The 4S Web-Marketing Mix model". **Electronic Commerce Research and Applications**. 1(1): 57-76

CORSTJENS, J. AND CORSTJENS, M., 1999. **Store Wars, the Battle for Mindspace and Shelfspace**. Wiley: New York

CRESWELL, J.W., 1994. **Research Design: Qualitative and Quantitative Approaches**. Sage Publications: London

CRITTENDEN, V.L., 2005. "The Rebuilt Marketing Machine". **Business Horizons**. 48(5): 409-420

CROMWELL, J.B., LABYS, W.C. AND TERRAZA, M., 1994. **Univariate Tests for Time Series Models**. Sage Publications: London

CRYER, J.D., 1986. **Time Series Analysis**. PWS Publishers: Boston

DARBELLAY, G.A. AND SLAMA, M., 2000. "Forecasting the Short-Term Demand for Electricity: Do Neural Networks Stand a Better Chance?". **International Journal of Forecasting.** 16(1): 71-83

DATTA, Y., 1996. "Market Segmentation:An Integrated Framework". **Long Range Planning**. 29(6): 797-811

DE GOOIJER, J.G. AND HYNDMAN, R.J., 2006. "25 Years of Time Series Forecasting". **International Journal of Forecasting**. 22(3): 443- 473

DENTON, S., 2005. **Measuring SA's 'Second' Economy**. [On-line] http://www.southafrica.info/doing_business/trends/newbusiness/stats-sa-121005.hlm (accessed 20 October 2006)

DEVEY, R., SKINNER, C. AND VALODIA, I., 2006 . **Second Best? Trends and Linkages in the Informal Economy in South Africa**. Unpublished [Available online] http://www.sarpn.org.za/documents/d0002003/Economy_SA_Feb2006.pdf (accessed 20 October 2006)

DIEBOLD, F.X. AND KILIAN, L., 1999 . **Unit Root Tests are Useful for Selecting Forecasting Models**. Unpublished [Available online] http://www.ssc.upenn.edu/~diebold/ (accessed 26 September 2006)

DUFFY, D.L., 2005. "Direct Selling as the Next Channel". **Journal of Consumer Marketing.** 22(1): 43-45

ELSKEN, T., 1996. "Even on Finite Test Sets Smaller Nets may Perform Better". **Neural Networks**. 10(2): 369-385

ENDERS, W., 2004. **Applied Econometric Time Series**. John Wiley and Sons: New Jersey

ERDEM, T., SWAIT, J. AND LOUVIERE, J., 2002. "The impact of brand credibility on consumer price sensitivity". **International Journal of Research in Marketing**. 1(19): 1-19

FERRELL, O.C. AND HARTLINE, M.D., 2005. **Marketing Strategy (3e)**. Thomson South-Western: New York

FUJITA, O., 1998. "Statistical Estimation of the Number of Hidden Units for Feed Forward Neural Networks". **Neural Networks.** 11(5): 851-859

FUNAHASHI, F., 1989. "On the approximate realization of continuous mappings by neural networks". **Neural Networks.** 2(3): 183-192

GARDNER, E.S. AND MCKENZIE, E., 1985. "Forecasting Trends in Time Series". **Management Science**. 31(10): 1237-1246

GARDNER, E.S. AND MCKENZIE, E., 1988. "Model Identification In Exponential Smoothing". **Journal of the Operational Research Society**. 39(9): 863-867

GARDNER, E.S. AND MCKENZIE, E., 1989. "Exponential Smoothing with Damped Trends". **Management Science**. 35(3): 372-376

GARDNER, E.S., 1990. "Evaluating Forecast Performance in Inventory Control". **Management Science**. 36(4): 490-499

GOLDSMITH, R.E, 1999. "The Personalised Marketplace - Beyond the 4P". **Marketing Intelligence and Planning**. 17(4): 178-185

GREWAL, D. AND MARMORSTEIN, H., 1994. "Market Price Variation, Perceived Price Variation, and Consumers' Price Search Decisions for Durable Goods". **The Journal of Consumer Research.** 21(3): 453-460

GRUENWALD, G., 1997. **How to Create Profitable New Products**. NTC Business Books: Chicago

GUJARATI, D.N., 2006. **Essentials of Econometrics. (International Edition)**. McGraw Hill: Boston

GUPTA, M.C. AND BENEDETTO, C.A., 2006 . "Optimal pricing and advertising strategy for introducing a new business product with threat of competitive entry" **Industrial Marketing Management**. Unpublished [Available online] http://www.sciencedirect.com/science?_ob=MImg&_imagekey=B6V69-4JKYWGX-1-1S&_cdi=5809&_user=736737&_orig=search&_coverDate=03%2F31%2F2006&_sk=999999999&view=c&_alid=438152073&_rdoc=1&wchp=dGLbVzb-zSkzS&md5=c2b448a25d6cd03fc4242630fdf3bade&ie=/sdarticle.pd (accessed 20 August 2006)

HAKANSSON, H. AND WALUSZEWSKI, A., 2005. "Developing a new Understanding of Market: Reinterpreting the 4P's". **Journal of Business and Industrial Marketing**. 20(3): 110-117

HAMID, S.A., 2004. **Primer on Using Neural Networks For Forecasting Market Variables.** [On-line] http://www.snhu.edu/files/pdfs/NeuralNetworks2004-03.pdf#search=%22neural%20networks%20as%20forecasting%20experts%22 (accessed 10 September 2006)

HART, C., 2005. **Doing your Masters Dissertation.** Sage Publications: London

HART, S AND TZOKAS, E., 2000. "New Product Launch "Mix" in Growth and Mature Product Markets". **Benchmarking: An International Journal.** 7(5): 389-405

HARVEY, A.C., 1994. **Time Series Models.** MIT Press: Cambridge

HILL, T., O'CONNER, M. AND REMUS. W., 1996. "Neural Network Models for Time Series Forecasts". **Management Science.** 42(7): 1082-1092

HOLM, O., 2006. "Integrated Marketing Communications: From Tactics to Strategy". **Corporate Communications: An International Journal.** 11(1): 23-33

HUGHES, J. AND SHARROCK, W., 1997. **The Philosophy of Social Research.** Longman: London

HULBERT, J.M., CAPON, N., AND PIERCY, N.F., 2003. **Total Integrated Marketing, Breaking the Bounds of the Function.** Kogan Page: London

JEUNET, J., 2006. "Demand Forecast Accuracy and Performance of Inventory Policies Under Multi-level Rolling Schedule Environments". **International Journal of Production Economics.** 10(3): 401-419

JOHNSON AND JOHNSON, 2006. **Our Company.** [On-line] http://www.jnj.com/our_company/history/history_section_4.htm (accessed 12 September 2006)

JOHNSON, R.A. AND WICHERN, D.A., 2002. **Applied Multivariate Statistical Analysis (5e).** Pearson Education: New Jersey

JOHNSTON, J. AND DINARDO, J., 1997. **Econometric Methods.** McGraw Hill: Singapore

JOLSON, M.A. AND HISE, R.T., 1973. **Quantitative Techniques for Marketing Decisions.** McMillan Company: New York

JÜTTNER, U., GODSELL, J. CHRISTOPHER, M.G, 2006. "Demand Chain Alignment

Competence - Delivering Value through Product Life Cycle Management". **Industrial Marketing Management**. 8(35): 989-1001

KALWANI, M.U. AND YIM, C.K., 1992. "Consumer Price and Promotion Expectations: An Experimental Study". **Journal of Marketing Research**. 29(1): 90-100

KELLER, K.L., 1987. "Memory Effects in Advertising: The Effect of Advertising Retrieval Cues on Brand Evaluations". **The Journal of Consumer Research**. 14(3): 316-333

KELLER, K.L., 1998. **Strategic Brand Management**. Prentice Hall: Upper Saddle River

KENDALL, M. AND ORD, K.J., 1990. **Time Series**. Edward Arnold: Suffolk

KITCHEN, P.J., 1989. "Competitiveness in FMCG Markets". **European Journal of Marketing**. 23(1): 41-51

KOTLER, P., 1972. "A Generic Concept of Marketing". **Journal of Marketing**. 36(2): 46-54

KOTLER, P., 1994. **Marketing Management (8e)**. Prentice Hall: New Jersey

KOLTER, P., 1998. **Marketing Management**. Prentice Hall: Englewood Cliffs

KRAUTER, J., 1999. "Inventory Theory: New Perspectives for Corporate Management". **International Journal of Production Economics**. 1-3(59): 129-134

KRISHNAMURTHI, L. AND RAJ, S.P., 1991. "An Empirical Analysis of the Relationship between of Brand Loyalty and Consumer Price Elasticity". **Marketing Science**. 10(2): 172-183

KUO, R.J. AND XUE, K.C., 1998. "A Decision Support System for Sales Forecasting through Fuzzy Neural Networks with Asymmetric Fuzzy Weights". **Decision Support Systems**. 24(2): 105-126

LAMB, C.W., HAIR, J.F. AND MCDANIEL, C., 2002. **Marketing (6e)**. South-Western: Cincinnati

LANCIONI, R.A., 1988. "The Importance of Price in International Business Development". **Asia Pacific International Journal of Business Logistics**. 23(11): 45-50

LANCIONI, R.A., 2004. "A Strategic Approach to Industrial Product Pricing: The Pricing Plan". **Industrial Marketing Management**. 34(2): 177-183

LEE, A., 1991. "Integrating Positivist and Interpretative Approaches to Organizational Research". **Organisational Science**. 2(4): 342-365

LEONE, R.P. AND SRINIVASAN,, 1996. "Coupon Face Value: Its Impact on Coupon Redemption, Brand Sales, and Brand Profitability". **Journal of Retailing**. 73(3): 273-289

LEUTHOLD, R.M., 1975. "On The Use of Theils Inequality Coefficients". **American Journal of Agricultural Economics**. 57(2): 344-346

LEVEDAHL, J.W., 1984. "Marketing, Price Discrimination, and Welfare: A Comment". **Southern Economic Journal**. 50(3): 886-891

LEWIS, C.D., 1975. **Demand Analysis and Control**. Lexington Books: Westmead

LICHTENSTEIN, D.R., RIDGEWAY, N.M. AND NETEMEYER, R.G., 1993. "Price Perceptions and Consumer Shopping Behaviour: A Field Study". **Journal of Marketing Research**. 30(2): 234-245

MADDALA, K., 2003. **Introduction to Econometrics**. John Wiley and Sons: Chichester

MAGRATH, A.J., 1995. **Marketing Strategies for Growth in Uncertain Times**. NTC Publishing Group: Illinois

MAKRIDAKIS, S., 1978. "Time Series Analysis and Forecasting: An Update and Evaluation". **International Statistical Review**. 46(3): 255-278

MAXWELL, S., 1998. "Pricing education in the United States of America: Responding to the Needs of a Business". **Journal of Product and Brand Management**. 7(4): 336-341

MELA, C.F., GUPTA, S. AND JEDIDI,K., 1998. "Assessing long-term promotional influences on market structure". **International Journal of Research in Marketing**. 15(2): 89-107

MILLER, T. AND LIBERATORE, M., 1993. "Seasonal Exponential Smoothing with Damped Trends: An Application for Production Planning". **International Journal of Forecasting**. 9(4): 509-515

MILLS, T.C., 1990. **Time Series Techniques for Economists**. Cambridge University Press: Cambridge

MINGERS, J., 2001. "Combining IS Research Methods: Towards a Pluralist Methodology".

Information Systems Research. 12(3): 240-259

MONROE, K.B., 1973. "Buyers' Subjective Perceptions of Price". **Journal of Marketing Research**. 10(1): 70-80

MORRELL, A.J.H., 1967. **Problems of Stock and Storage**. Oliver and Boyd: Edinburgh

MORRIS, R., 1989. **The Marketing Principle**. Butterworths: Durban

MULHERN, F.J. AND LEONE, R.P., 1995. "Measuring Market Response to Price Changes: A Classification Approach". **Journal of Business Research**. 3(33): 197-205

NAM, K. AND SCHAEFER, T., 1995. "Forecasting international airline passenger traffic using neural networks". **The Logistics and Transportation Review**. 31(3): 239-251

NAUCK, D., KLAWONN, F. AND KRUSE, R., 1997. **Foundations of Neuro-Fuzzy Systems**. John Wiley & Sons: Midsomer Norton

O'MALLEY, L., PATTERSON, M. AND EVANS, M., 1999. **Exploring Direct Marketing (5e)**. Thompson Learning: Surrey

PARKIN, M., POWELL, M. AND MATTHEWS, K., 2000. **Economics (5e)**. Addison-Wesley: Madrid

PATTERSON, D.W., 1990. **Introduction to Artificial Intelligence and Expert Systems**. Prentice-Hall International: London

PEGELS, C.C., 1969. "Exponential Forecasting: Some New Variations". **Management Science**. 15(5): 311-415

PICK 'N PAY, 2006. **About Us**. [On-line] http://www.pnp.co.za/aboutus/st_targetmarket.html (accessed 12 September 2006)

RAFIQ, M. AND AHMED, P.K., 1995. "Using the 7Ps as a Generic Marketing Mix: an Exploratory Survey of UK and European Marketing Academics". **Marketing Intelligence and Planning**. 13(9): 4-15

REED, R.D. AND MARKS, R.J., 1999. **Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks**. MIT Press: London

ROBERTS, S.J. AND PENNY, W., 1997. "Neural Networks: Friends or Foes". **Sensor Review**. 17(1): 64-70

SAKIA, R.M., 1992. "The Box-Cox Transformation: A Review". **The Statistician**. 41(2): 169-178

SCHITTENKOPF, C., DECO, G. AND BRAUER, W., 1997. "Two Strategies to Avoid Overfitting in Feedforward Networks". **Neural Networks**. 10(3): 505-516

SHANKER, M., HU, M.Y. AND HUNG, M. S., 1996. "Effect of data standardization on neural network training". **Omega**. 24(4): 385-397

SHIPLEY, D. AND JOBBER, D., 2001. "Integrative Pricing via the Pricing Wheel". **Industrial Marketing Management**. 3(30): 301-314

SHUMWAY, R.H., 1988. **Applied Statistical Time Series Analysis**. Prentice Hall: Englewood Cliffs

SIMA, J., 1996. "Back-Propagation in Not Efficient". **Neural Networks**. 9(6): 1017-1023

SIMKIN, L., 2000. "Marketing is Marketing, Maybe!". **Marketing Intelligence and Planning**. 18(3): 154-158

SON, M., HAHN, M. AND KANG, H., 2006 . "Why firms do co-promotions in mature markets?" **Journal of Business Research**. Unpublished [Available online] http://www.sciencedirect.com/science?_ob=MImg&_imagekey=B6V7S-4K7WJ7Y-1-S&_cdi=5850&_user=736737&_orig=search&_coverDate=06%2F23%2F2006&_sk=999999999&view=c&_alid=438149620&_rdoc=1&wchp=dGLbVlb-zSkzS&md5=b3e21cf45ea23785b3c2e555cd5a0816&ie=/sdarticle.pdf (accessed 20 August 2006)

SRINIVASAN, S., POPKOWSKI LESZCZYC, P.T.C., BASS, F.M., 2000. "Market share response and competitive interaction: The impact of temporary,". **International Journal of Research in Marketing**. 17(4): 281-305

SUDARSAN, R., FENVES, S.J., SRIRAM, R.D. AND WANG, F., 2005. "A product information modelling framework for product life cycle management". **Computer-Aided Design**. 13(37): 1399-1411

SVENSSON,G., 2002. "Supply Chain Management: the Reintegration of Marketing Issues in Logistics Theory and Practise". **European Business Review**. 14(6): 426-436

TAIO, G.C., 2001. **Time Series: ARIMA Methods**. [On-line] http://www.sciencedirect.com/science?_ob=MImg&_imagekey=B6WVS-4DN8TNP-3N6-X&_cdi=23486&_user=736737&_orig=search&_coverDate=10%2F28%2F2004&_sk=99999&view=c&wchp=dGLzVlz-zSkWz&md5=58b05d05ba109ad59fd742d983958e4d&ie=/sdarticle.pdf (accessed 14 September 2006)

TAIO, G.C. AND BOX, G.E.P., 1981. "Modelling Multiple Time Series with Applications". **Journal of the American Statistical Association**. 76(376): 802-816

TAKADA, H. AND BASS, F.M., 1998. "Multiple Time Series Analysis of Competitive Marketing Behaviour". **Journal of Business Research**. 43(2): 97-107

TANG, Z. AND FISHWICK, P.A., 1993. **Feed-Forward Neural Nets as Models for Time Series Forecasting**. [On-line] http://http://www.cise.ufl.edu/~fishwick/tr/tr91-008.html (accessed 10 September 2006)

TAYLOR, J.W., 2003. "Exponential smoothing with a damped multiplicative trend". **International Journal of Forecasting**. 19(4): 715-725

TERRE BLANCHE, M. AND DURRHEIM, K., 1999. **Research in Practise**. University Of Cape Town Press: Cape Town

THOMAS, A.B., 1968. **Stock Control in Manufacturing**. Gower Press: Aylesbury

TON SU, C. AND LIN HSIEH, K., 1997. "Applying Neural Network Approach to Achieve Robust Design for Dynamic Quality Characteristics". **International Journal of Quality & Reliability Management**. 10(5): 509-519

TULL, D.S., 1965. "The Carry-Over Effect of Advertising". **Journal of Marketing**. 29(2): 46-53

VAN WEZEL, M.C. AND BAETS, W.R.J., 1995. "Predicting Market Response With a Neural Network: The Case of Fast Moving Consumer Goods". **Marketing Intelligence and Planning**. 13(7): 23-30

VENUGOPAL, V. AND BAETS, W., 1994. "Neural Networks and Statistical Techniques in Marketing Research". **Marketing Intelligence and Planning**. 12(7): 30-38

WARNER, B. AND MISRA, M., 1996. "Understanding Neural Networks as Statistical Tools". **The American Statistician**. 50(4): 284-293

WEBER, J.A. AND DHOLAKIA, U., 1998. "Planning Market Growth in Mature Business Markets". **Industrial Marketing Management**. 27(5): 401-428

WINTERS, P.R., 1960. "Forecasting Sales by Exponentially Weighted Moving Averages". **Management Science**. 6(3): 324-342

WONG, B.K., LAI, V.S. AND LAM, J., 2000. "A Biography of Neural Network Business Applications Research: 1994-1998". **Computers and Operations Research**. 11-12(27): 1045-1076

ZENOR, M.J., BRONNENBURG, B.J., AND MCALISTER, L., 1998. "The Impact of Marketing Policy on Promotional Price Elasticities and Baseline Sales". **Journal of Retailing and Public Services**. 5(1): 25-32

ZHANG, G., EDDY PATUWO, B. AND HU, M.Y., 1998. "Forecasting with Artificial Neural Networks:The State of the Art". **International Journal of Forecasting**. 14(1): 35-62

APPENDICES

Appendix A

Script of Neural Network Program

```
%David Winn - 2006

%Clear all the data which may be stored in memory
clear all
close all
clc

echo on
% A Neural Network to predict the demand for a specific deodorant over a various
number of months.
%
%
% Developed in Partial Fulfillment of a Masters Thesis at Rhodes
% Univeristy, under the superivsion of:
% Prof S.Radloff
% Dr. R.Elliott
% Prof M.Burton
%
%
% Completed By
% David Winn
% Student Number - G02W1053
% For additional information, please contact Dave.winn@gmail.com
%
echo off
disp('-------------------------------------')
R=xlsread('C:\...\male\M1.xls');


%Deodorants would be expected to be sold in a cyclical pattern, with
%greater use in summer when it is warmer. As such, a sin and cos variable has
been created to take into account any cyclical nature.

%%%%%%%%%%%%%%%%%%%%
%Data Preparation  %
%%%%%%%%%%%%%%%%%%%%

%Set month as sin and cos variables, to denote a position on a circle, thus
%ensuring that each similarly named month is recognised as equal, e.g.
%January for every year, is recognised as being a January, and the
%identical values prescribed.

R=R';
r=size(R);
r1=r(1);
r2=r(2);
a = rem(r1,12);
b=floor(r1/12);
q1=repmat([1:12],1,b)';
q2=[q1;(1:a)'];
for i=1:r1
    R(i,(r2+1))= cos(2*pi*q2(i,1)/12);
```

```
    R(i,(r2+2))= sin(2*pi*q2(i,1)/12);
end

%Setting an inflation variable if required
% infl=0.036;
% R(:,15) = diag(R(:,1)*(1+infl/12).^(r1-(1:r1)));
% R(:,13) = diag(R(:,2)*(1+infl/12).^(r1-(1:r1)));
% R(:,14) = diag(R(:,8)*(1+infl/12).^(r1-(1:r1)));
 h = 3 ;     %number of months memory occurs for

%Brand Retention
m1 = 0.4;
m2 = 0.6;

%inputs
if  find (R(:,6)==0);                       %Determining when the product was
introduced, and only calculating the demand from then to prevent skewed data
    no_product = find (R(:,6)==0);
    no_product = max(no_product);
    Product_exists = no_product+1;

    p1= R((Product_exists:end),end-1);      %cos variable of month
    p2= R((Product_exists:end),end);        %sin variable of month
    p3= R((Product_exists:end),3);          %promotion spend
    p4= R((Product_exists:end),4);          %avg price
    p5= R((Product_exists:end),3);          %Number of stores sold in
    p6= R((Product_exists:end),5);          %Products in category
    p7= R((Product_exists:end),6);          %Number of products in line
    p8= R((Product_exists:end),7);          %Number of brands
    p9=R((Product_exists:end),14);          %Industry Price
    p10= R((Product_exists:end),9);         %Total market
    p11= R((Product_exists:end),10);        %total marketing


    %Adjust the promotion spend to reflect a percentage of previous months, as
    %people remember promotions even once it is over. This is estimated.
    r3 = size(p3);
    r3=r3(1);
    p12=p3
    for i=h:r3;
            p12(i) = m1*p3(i-2) + m2*p3(i-1)+ p3(i) ;
    end

% Inputs
    p=[p1 p2 p3 p4 p5 p6 p7 p8 p9 p10 p12']';
%targets
    t=R((Product_exists:end),11)';  %products sold
else %In the event that there is a full data range
    p1= R(:,end-1);     %cos variable of month
    p2= R(:,end);       %sin variable of month
    p3= R(:,3);         %promotion spend
    p4= R(:,4);         %avg price
    p5= R(:,3);         %Number of stores sold in
    p6= R(:,5);         %Products in category
    p7= R(:,6);         %Number of products in line
    p8= R(:,7);         %Number of brands
    p9=R(:,10);         %Industry Price
    p10= R(:,9);        %Total market
    p11= R(:,10);       %total marketing

    %Adjust the promotion spend to reflect a percentage of previous months, as
```

```
    %people remember promotions even once it is over. This is estimated.
    r3 = size(p3);
    r3=r3(1);
    for i=h:r3;
            p12(i) = m1*p3(i-2) + m2*p3(i-1)+ p3(i) ;
    end
    p=[p1 p2 p3 p4 p5 p6 p7 p8 p9 p10 p12']';
    %targets
    t=R(:,11)';  %products sold
end


%Normalise the data to put all the data in the same range. This speeds up
%the training process, but does not necessarily improve the results
[pn mp Mp tn mt Mt] = premnmx(p,t);

targets_actual = [];        %Define variables for actual targets
predicted_total = [];       %Define variables for predicted totals
% y= 100;                    %Number of simulations
y = input('The number of simulations =  ');
n = input('The number of months back you wish to simulate from to check the
system =  ');         %number of months back to start prediction back
% n = 6;

i=1:n;      % Setting the loop of months
i=n-i+1;    % Reversing the sequence to calculate backwards
for i=i
    train_data = pn(:,1:end-i);
    train_targets=tn(:,1:end-i);
    test_data = pn(:,end-i+1:end-i+1);
    test_targets = tn(:,end-i+1:end-i+1);
    [total,test_targets_actual,estimatenet] =
demandestimatefunction(test_data,test_targets,train_data,train_targets,pn,mp,Mp,
tn,mt,Mt,y); %Please refer to the relevant function for details shown below
    predicted_total  = [predicted_total, total];
    targets_actual = [targets_actual,test_targets_actual ];
end



%%%%%%%%%%%%%%%%%%%%
%Displaying Results%
%%%%%%%%%%%%%%%%%%%%
title = (n-(1:n))';
format short g

Actual_error = (mean(predicted_total)-targets_actual);
Percentage_error =  Actual_error./targets_actual;
out_demand = [title,mean(predicted_total)'];
out_error = [title,round(Actual_error)'];
out_percentage = [title,(Percentage_error*100)'];

disp('Results')
disp('---------------------------')
disp('Predicted demand')
fprintf('Predicted demand for %1.0f months ago was %2.0f units\n',out_demand')
disp('---------------------------')
disp('Actual Error')
fprintf('Actual error for %1.0f months ago is %2.0f units\n',out_error')
disp('---------------------------')
disp('Percentage Error')
```

```
fprintf('Percentage error for %1.0f months ago is %1.3f%%\n',out_percentage')
disp('--------------------------')
```

```
%%%%%%%%%%%%%%%%%%%
% Plot of the data%
%%%%%%%%%%%%%%%%%%%
close all;
for i=1:n;
    month = n-i;
    figure;
    hold on;
    line1 = linspace(1,y,501);
    line2 = linspace(1,y,501);
    plot(predicted_total(:,i),'r-');
%     ylim([0,max(Atotal(:,i))])              %To scale from 0 - max value if
required
    plot(line1,targets_actual(:,i),'b:');
    plot(line2,mean(predicted_total(:,i)),'g-');
    xlabel('Simulation number');
    ylabel('Predicted Demand');
    TITLE(sprintf('Graph of Demand Predictions for %1.0f month ago',month))
    legend('predictions','avg prediction','actual','location','best')
end
```

Neural Network Function

```
function [total,test_targets_actual,estimatenet] =
demandestimatefunction(test_data,test_targets,train_data,train_targets,pn,mp,Mp,
tn,mt,Mt,y)

%David Winn - Function file
% A Neural Network to predict the deodorant demand for various months.
%
%
% Developed as Partial Fulfilment of a Masters Thesis as Rhodes
% University, under the supervision of Prof M.Burton, Prof S.Radloff, Dr.
% R.Elliott
%
% Complete By
% David Winn
% Student Number - G02W1053
% For additional Information, please contact Dave.winn@gmail.com

count=0;
total = [];
actual=[];

while count<y       %repeat for y simulations to find an average to make it more
accurate
    count=count+1;

    %si = number of neurons
    s1=17;
```

```
    s2=4;
    s3=1;

    %training method
    tm='trainscg';

    net=newff(minmax(pn),[s1,s2,s3], {'tansig', 'tansig','purelin'}, tm);

    %Net training
    net.TrainParam.epochs = 500;
    net.trainparam.goal=0.005;
    net.trainparam.show= 500;

    %initiate the weights and biases
    net=init(net);


    %train the net on the training data
    [net,tr] = train(net,train_data,train_targets);
    estimatenet=net;

%%%%%%%%%%%%%
%Simulations%
%%%%%%%%%%%%%

    %simulate with the normalised data
    adata=sim(estimatenet,pn);                 %Simulation on all the data to check
it has run corretly
    atrain=sim(estimatenet,train_data);        %Simulation on all the training data
to check it has run corretly
    atest=sim(estimatenet,test_data);          %Simulation on all the test data to
make the predictions

    %postprocess the training data back into its original scale
    adata=postmnmx(adata,mt,Mt);
    atest = postmnmx(atest,mt,Mt);
    atrain = postmnmx(atrain,mt,Mt);
    atargets = postmnmx(tn,mt,Mt);

    %post process the input data  back into its original scale
    train_targets_actual=postmnmx(train_targets,mt,Mt);
    test_targets_actual = postmnmx(test_targets,mt,Mt);
    total= [total;atest];
end
```

Appendix B

Male Results of Forecast 1 to 6.

Product		Actual	Neural Network	% Diff	Time Series	% Diff	Previous Month	% Diff
M1	1	2,348	2,184	-6.98	2,963	26.19	2,798	19.17
	2	2,177	2,356	8.22	2,439	12.03	2,348	7.85
	3	2,246	2,250	0.18	2,274	1.25	2,177	-3.07
	4	2,870	2,515	-12.37	2,346	-18.26	2,246	-21.74
	5	2,434	2,686	10.35	3,011	23.71	2,870	17.91
	6	3,184	2,832	-11.06	2,527	-20.63	2,434	-23.56
M2	1	2,116	2,255	6.57	2,136	0.95	2,162	2.17
	2	1,944	1,938	-0.31	2,093	7.66	2,116	8.85
	3	1,848	1,982	7.25	1,901	2.87	1,944	5.19
	4	1,822	1,972	8.23	1,792	-1.65	1,848	1.43
	5	1,615	1,813	12.26	1,962	21.49	1,822	12.82
	6	1,851	1,919	3.67	1,542	-16.69	1,615	-12.75
M3	1	28,309	21,718	-23.28	16,572	-41.46	17,080	-39.67
	2	22,789	27,303	19.81	29,663	30.16	28,309	24.22
	3	31,071	31,072	0.00	23,007	-25.95	22,789	-26.66
	4	37,107	32,991	-11.09	32,285	-12.99	31,071	-16.27
	5	31,247	37,294	19.35	39,127	25.22	37,107	18.75
	6	38,588	33,674	-12.73	32,492	-15.80	31,247	-19.02
M4	1	520	521	0.19	501	-3.65	554	6.54
	2	472	493	4.45	471	-0.21	520	10.17
	3	523	501	-4.21	421	-19.50	472	-9.75
	4	555	652	17.48	484	-12.79	523	-5.77
	5	495	594	20.00	552	11.52	555	12.12
	6	590	537	-8.98	225	-61.86	495	-16.10

Product		Actual	Neural Network	% Diff	Time Series	% Diff	Previous Month	% Diff
M6	1	2,684	2,865	6.74	2,543	-5.25	2,512	-6.41
	2	2,056	2,438	18.58	2,733	32.93	2,684	30.54
	3	2,131	2,041	-4.22	2,020	-5.21	2,056	-3.52
	4	2,348	2,377	1.24	2,111	-10.09	2,131	-9.24
	5	2,384	2,412	1.17	2,358	-1.09	2,348	-1.51
	6	2,930	2,762	-5.73	2,407	-17.85	2,384	-18.63
M7	1	5,024	4,943	-1.61	5,121	1.93	4,988	-0.72
	2	4,437	4,541	2.34	5,123	15.46	5,024	13.23
	3	5,974	5,396	-9.68	4,478	-25.04	4,437	-25.73
	4	6,003	6,442	7.31	6,219	3.60	5,974	-0.48
	5	5,684	6,274	10.38	6,214	9.32	6,003	5.61
	6	6,152	6,000	-2.47	5,887	-4.31	5,684	-7.61
M9	1	9,450	9,765	3.33	10,136	7.26	9,642	2.03
	2	8,724	9,626	10.34	8,606	-1.35	9,450	8.32
	3	10,119	11,393	12.59	8,407	-16.92	8,724	-13.79
	4	11,103	10,823	-2.52	10,032	-9.65	10,119	-8.86
	5	9,279	10,343	11.47	11,496	23.89	11,103	19.66
	6	10,168	9,805	-3.57	8,685	-14.58	9,279	-8.74
M11	1	27,801	27,283	-1.86	30,298	8.98	24,394	-12.25
	2	27,342	26,459	-3.23	29,276	7.07	27,801	1.68
	3	28,784	32,814	14.00	24,460	-15.02	27,342	-5.01
	4	36,866	33,866	-8.14	23,763	-35.54	28,784	-21.92
	5	38,960	36,468	-6.40	33,618	-13.71	36,866	-5.37
	6	48,214	46,023	-4.54	31,759	-34.13	38,960	-19.19
M12	1	3,533	3,706	4.90	3,832	8.46	3,437	-2.72
	2	4,561	4,701	3.07	4,000	-12.30	3,533	-22.54
	3	4,937	6,110	23.76	4,598	-6.87	4,561	-7.62
	4	4,781	5,290	10.65	4,733	-1.00	4,937	3.26
	5	6,321	6,147	-2.75	5,389	-14.74	4,781	-24.36
	6	6,474	6,442	-0.49	5,307	-18.03	6,321	-2.36

Product		Actual	Neural Network	% Diff	Time Series	% Diff	Previous Month	% Diff
M13	1	5,244	6,191	18.06	4,556	-13.12	4,945	-5.70
	2	5,853	5,771	-1.40	4,790	-18.16	5,244	-10.40
	3	6,173	8,013	29.81	4,741	-23.20	5,853	-5.18
	4	6,905	8,914	29.09	4,538	-34.28	6,173	-10.60
	5	7,917	7,690	-2.87	5,427	-31.45	6,905	-12.78
	6	12,614	10,707	-15.12	5,416	-57.06	7,917	-37.24
M14	1	626	633	1.12	655	4.63	637	1.76
	2	582	653	12.20	616	5.84	626	7.56
	3	515	659	27.96	484	-6.02	582	13.01
	4	634	984	55.21	439	-30.76	515	-18.77
	5	924	1,019	10.28	710	-23.16	634	-31.39
	6	1,193	1,147	-3.86	1,091	-8.55	924	-22.55
M15	1	4,910	4,808	-2.08	6,113	24.50	7,989	62.71
	2	4,385	4,708	7.37	3,084	-29.67	4,910	11.97
	3	4,288	4,390	2.38	2,597	-39.44	4,385	2.26
	4	6,560	8,012	22.13	2,435	-62.88	4,288	-34.63
	5	4,514	5,062	12.14	4,713	4.41	6,560	45.33
	6	5,511	3,652	-33.73	2,763	-49.86	4,514	-18.09

Female Results of Forecast 1 to 6.

Product		Actual	Neural Network	% Diff	Time Series	% Diff	Previous Month	% Diff
F1	1	1,921	1,804	-6.09	1,671	-13.01	1,813	-5.622
	2	2,186	2,374	8.60	1,790	-18.12	1,921	-12.123
	3	2,401	2,059	-14.24	2,073	-13.66	2,186	-8.955
	4	2,390	2,602	8.87	2,306	-3.51	2,401	0.460
	5	2,040	2,377	16.52	2,299	12.70	2,390	17.157
	6	2,120	2,132	0.57	1,927	-9.10	2,040	-3.774
F3	1	10,276	10,298	0.21	8,445	-17.82	9,683	-5.771
	2	9,129	10,060	10.20	9,077	-0.57	10,276	12.564
	3	10,768	10,703	-0.60	7,955	-26.12	9,129	-15.221
	4	14,457	13,322	-7.85	13,501	-6.61	10,768	-25.517
	5	15,503	15,296	-1.34	12,907	-16.75	14,457	-6.747
	6	17,125	15,348	-10.38	13,901	-18.83	15,503	-9.472
F4	1	3,009	3,417	13.56	2,890	-3.95	3,201	6.381
	2	2,919	3,229	10.62	2,789	-4.45	3,009	3.083
	3	2,957	2,916	-1.39	2,688	-9.10	2,919	-1.285
	4	2,729	3,233	18.47	2,750	0.77	2,957	8.355
	5	2,200	3,361	52.77	2,515	14.32	2,729	24.045
	6	2,188	3,066	40.13	1,928	-11.88	2,200	0.548
F5	1	67	72	7.46	125	86.57	141	110.448
	2	30	47	56.67	33	10.00	67	123.333
	3	8	10	25.00	0	-100.00	30	275.000
F6	1	2,565	2,469	-3.74	2,320	-9.55	2,385	-7.018
	2	2,040	2,208	8.24	2,603	27.60	2,565	25.735
	3	2,099	2,097	-0.10	1,845	-12.10	2,040	-2.811
	4	2,285	2,392	4.68	1,848	-19.12	2,099	-8.140
	5	2,198	2,331	6.05	2,229	1.41	2,285	3.958
	6	2,516	2,405	-4.41	1,935	-23.09	2,198	-12.639

Product		Actual	Neural Network	% Diff	Time Series	% Diff	Previous Month	% Diff
F7	1	2,834	3,059	7.94	3,329	17.47	3,012	6.281
	2	2,349	2,652	12.90	3,177	35.25	2,834	20.647
	3	3,085	3,091	0.19	2,501	-18.93	2,349	-23.857
	4	3,477	3,784	8.83	2,969	-14.61	3,085	-11.274
	5	3,324	3,977	19.65	3,492	5.05	3,477	4.603
	6	3,418	3,702	8.31	2,929	-14.31	3,324	-2.750
F9	1	5,748	5,589	-2.77	4,980	-13.36	5,457	-5.063
	2	4,573	5,370	-88.26	5,637	23.27	5,748	25.694
	3	3,895	5,281	35.58	4,658	19.59	4,573	17.407
	4	2,864	4,330	51.19	3,264	13.97	3,895	35.999
	5	3,585	4,096	14.25	1,902	-46.95	2,864	-20.112
	6	3,945	4,533	14.90	2,305	-41.57	3,585	-9.125
F10	1	12,358	11,373	-7.97	11,134	-9.90	10,907	-11.741
	2	11,296	11,480	1.63	11,615	2.82	12,358	9.402
	3	12,211	11,212	-8.18	10,289	-15.74	11,296	-7.493
	4	16,187	15,989	-1.22	10,889	-32.73	12,211	-24.563
	5	19,634	18,421	-6.18	15,312	-22.01	16,187	-17.556
	6	23,496	22,183	-5.59	15,677	-33.28	19,634	-16.437
F11	1	2,317	2,172	-6.26	2,003	-13.55	2,013	-13.120
	2	2,149	2,367	10.14	2,205	2.61	2,317	7.818
	3	2,361	2,332	-1.23	2,050	-13.17	2,149	-8.979
	4	2,643	2,543	-3.78	2,343	-11.35	2,361	-10.670
	5	2,451	2,810	14.65	2,727	11.26	2,643	7.834
	6	2,645	2,474	-6.47	2,361	-10.74	2,451	-7.335
F12	1	7,098	6,431	-9.40	6,595	-7.09	7,033	-0.916
	2	7,148	6,862	-4.00	6,678	-6.58	7,098	-0.699
	3	7,617	8,051	5.70	6,832	-10.31	7,148	-6.157
	4	7,487	8,480	13.26	7,621	1.79	7,617	1.736
	5	9,369	8,555	-8.69	7,512	-19.82	7,487	-20.088
	6	12,684	10,510	-17.14	10,144	-20.03	9,369	-26.135

Product		Actual	Neural Network	% Diff	Time Series	% Diff	Previous Month	% Diff
F13	1	28,513	28,119	-1.38	26,621	-6.64	24,546	-13.913
	2	28,307	28,925	2.18	28,225	-0.29	28,513	0.728
	3	29,406	33,362	13.45	26,591	-9.57	28,307	-3.737
	4	37,319	33,146	-11.18	27,151	-27.25	29,406	-21.204
	5	31,676	37,873	19.56	39,715	25.38	37,319	17.815
	6	44,260	38,637	-12.70	30,734	-30.56	31,676	-28.432
F14	1	15,179	15,607	2.82	18,430	21.42	14,721	-3.017
	2	19,078	20,142	5.58	19,559	2.52	15,179	-20.437
	3	21,909	25,980	18.58	21,863	-0.21	19,078	-12.922
	4	21,308	23,696	11.21	23,840	11.88	21,909	2.821
	5	27,543	27,999	1.66	25,808	-6.30	21,308	-22.637
	6	27,812	29,192	4.96	27,246	-2.04	27,543	-0.967

Appendix C

Differencing Test

Test Statistic: $t=\frac{\bar{D}-\delta}{s_d/\sqrt{n}}$ was used to test for differences between the actual sales value and forecasted value. The rejection region was specified as $t_{n-1}(\alpha/2)$ at the 0.05 level of significance, therefore for male sales forecasts the rejection region was $t_{71}(0.05/2)=1.666$, and the female sales forecast rejection region was $t_{69}(0.05/2)=1.667$. Results which fail to reject the null hypothesis are indicated with a *. Theils U-statistic is also presented for all product forecasts. A group rejection value of all the male forecasts calculated together is also presented, similarly for female forecasts.

	Forecast Method			
Product	**Neural Network**	**Time Series**	**Previous Month**	**Theil U**
M1	-0.6819*	0.2265*	-0.3029 *	2.2736
M2	3.8747	0.4346*	0.8062 *	2.5844
M3	-0.3943*	-0.8022*	-1.1889 *	1.2397
M4	0.9339	-1.3762*	-0.2354 *	2.6958
M6	0.7487*	-0.3686*	-0.4415 *	2.0336
M7	0.3102*	-0.1202*	-0.6346 *	2.2322
M9	1.6932*	-0.4020*	-0.1748 *	2.5629
M11	-0.8026*	-1.8402	-2.5164	1.6657
M12	1.5019*	-2.0528	-1.9179	2.2078
M13	0.7132*	-2.6025	-1.8460	1.3659
M14	1.8374	-1.8011	-1.4319	2.3332
M15	0.1736	-1.7880	0.5178 *	3.3674
Group	-0.0419*	-2.3383	-2.1465	

	Forecast Method			
Product	Neural Network	Time Series	Previous Month	Theil U
F1	0.4750*	-1.7291	-0.5704 *	2.5307
F3	-0.9568*	-3.8547	-1.9248	1.2264
F4	3.0831	-0.8269*	2.0308	3.5118
F5	2.0426	0.8489*	2.0212	-[15]
F6	0.6756*	-0.9254*	-0.1779 *	2.2925
F7	3.4850	-0.0613*	-0.3769 *	2.3753
F9	-0.0486*	-0.6252*	0.7649 *	2.5778
F10	-3.0566	-2.7662	-2.5530	1.3513
F11	0.2488*	-1.4217	-1.1469	2.5278
F12	-0.9341*	-2.4664	-1.6779	1.3863
F13	0.0521*	-1.0906	-1.2568 *	1.8412
F14	2.8629	0.8431*	-2.0443	2.0972
Group	0.4935 *	-2.6406	-2.9338	

15 Theil value of infinity as $\sum (y_t - y_{t-1}) = 0$, which when applied to Theils, results in an infinity value

Appendix D

Factor Analysis

All analyses were performed using 4 factors and a varimax rotation

Results: M1

```
Uniquenesses:
   Promotion           RSP   StoreCount      Products ProductTotal
BrandTotal
       0.887         0.487        0.086         0.020        0.005
0.634
    AvePrice  TotalMarket
       0.005        0.005

Loadings:
             Factor1 Factor2 Factor3 Factor4
Promotion                             0.323
RSP          -0.104   0.676   0.130   0.168
StoreCount    0.533   0.451   0.619   0.206
Products      0.734   0.215   0.516   0.360
ProductTotal  0.942  -0.155   0.289
BrandTotal    0.454   0.197           0.340
AvePrice      0.223   0.965
TotalMarket   0.275   0.115   0.944   0.123

               Factor1 Factor2 Factor3 Factor4
SS loadings      2.054   1.719   1.657   0.440
Proportion Var   0.257   0.215   0.207   0.055
Cumulative Var   0.257   0.472   0.679   0.734

Test of the hypothesis that 4 factors are sufficient.
The chi-square statistic is 2.12 on 2 degrees of freedom.
The p-value is 0.346
```

Results : M11

```
Varimax rotation
Uniquenesses:
   Promotion           RSP   StoreCount      Products ProductTotal
BrandTotal
       0.751         0.005        0.073         0.005        0.307
0.604
    AvePrice  TotalMarket
       0.245        0.114

Loadings:
             Factor1 Factor2 Factor3 Factor4
Promotion    -0.110  -0.192          -0.443
```

```
RSP             0.195   0.970   0.101
StoreCount      0.862   0.257   0.340
Products        0.762   0.210   0.606
ProductTotal    0.165   0.716   0.149   0.363
BrandTotal      0.204   0.176   0.561
AvePrice        0.119   0.769   0.294   0.251
TotalMarket     0.897   0.100           0.260

               Factor1 Factor2 Factor3 Factor4
SS loadings      2.260   2.233   0.925   0.478
Proportion Var   0.283   0.279   0.116   0.060
Cumulative Var   0.283   0.562   0.677   0.737

Test of the hypothesis that 4 factors are sufficient.
The chi-square statistic is 0.4 on 2 degrees of freedom.
The p-value is 0.818
```

Results: F1

```
Uniquenesses:
   Promotion          RSP   StoreCount       Products Producttotal
BrandTotal
       0.778        0.528        0.005          0.455        0.005
0.005
    AvgPrice  TotalMarket
       0.166        0.323

Loadings:
             Factor1 Factor2 Factor3 Factor4
Promotion            -0.162          -0.437
RSP          -0.519  -0.329   0.144  -0.271
StoreCount    0.852   0.178  -0.151   0.464
Products                      0.730
Producttotal  0.989
BrandTotal    0.860   0.383  -0.326
AvgPrice      0.113   0.870           0.252
TotalMarket   0.572          -0.380   0.452

               Factor1 Factor2 Factor3 Factor4
SS loadings      3.062   1.076   0.838   0.758
Proportion Var   0.383   0.134   0.105   0.095
Cumulative Var   0.383   0.517   0.622   0.717

Test of the hypothesis that 4 factors are sufficient.
The chi-square statistic is 5.38 on 2 degrees of freedom.
The p-value is 0.0678
```

Results: F11

```
Uniquenesses:
   Promotion          RSP   StoreCount       Products ProductTotal
BrandTotal
```

```
       0.584        0.507        0.089        0.456        0.005
0.005
    AvgPrice  TotalMarket
       0.005        0.145

Loadings:
             Factor1 Factor2 Factor3 Factor4
Promotion     0.102   0.229           0.595
RSP          -0.119          -0.688
StoreCount    0.811   0.285   0.322   0.261
Products      0.697                   0.230
ProductTotal  0.839           0.351  -0.409
BrandTotal    0.960   0.245           0.100
AvgPrice      0.190   0.936  -0.112   0.267
TotalMarket   0.597           0.540   0.454

               Factor1 Factor2 Factor3 Factor4
SS loadings      3.187   1.075   1.011   0.930
Proportion Var   0.398   0.134   0.126   0.116
Cumulative Var   0.398   0.533   0.659   0.775

Test of the hypothesis that 4 factors are sufficient.
The chi-square statistic is 3.67 on 2 degrees of freedom.
The p-value is 0.16
```

Appendix E

Individual Forecasts

Certain variables which are required in the described Neural Network are required to be forecasted in a practical setting as the data is unlikely to be available as it includes information that are only generated once all competitors have sold their products. This includes the total number of competing brand in the industry, the number of competing products, total market size in the number of products sold and the average industry price for deodorants. Using forecasted values in a forecaster lowers the overall accuracy, but it is felt that in this research the variables can be estimated accurately as to not have a major effect on the accuracy of the Neural Network forecaster. The estimation of the variables are described below.

Total Brands

Total brands competing in the industry can be assumed to be the same as the previous forecast as the number of competitors remains fairly constant over time as shown in figure E.1. As a forecasting method, the forecast for the following period can be assumed to be the same as the current period.

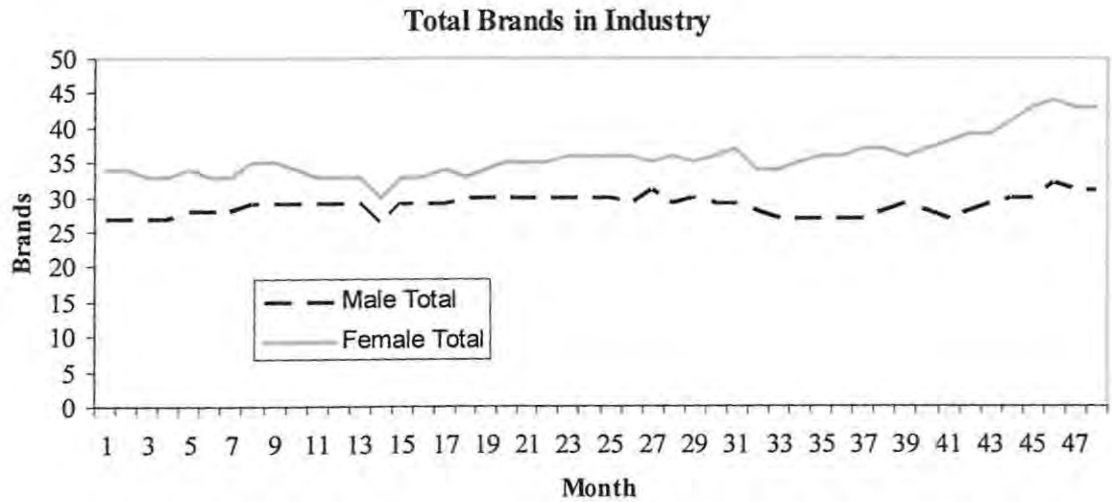


Figure E.1: Brand Total

Product Total

The number of competing products is constantly changing as firms introduce new variants and take others off the market and is shown in figure E.2. There does not appear to be any trends and so predictions are very difficult. If this information cannot be obtained through research, the best method of forecast may be to assume that the total products available will be the same as available in the previous month.

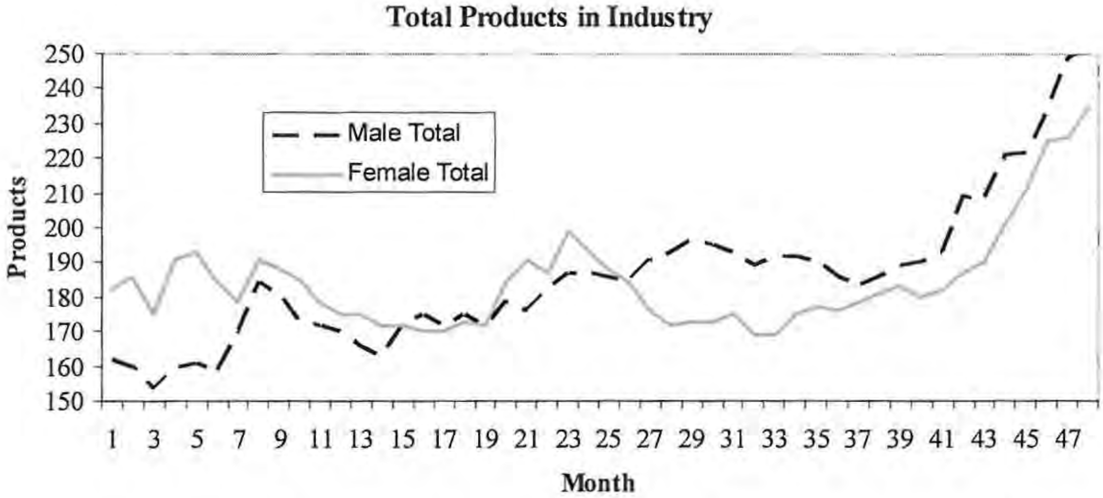


Figure E.2: Total Products in Industry

Total Units Sold

Total units sold displays a strong seasonal pattern. Forecasting this was complicated, and no form of ARMA model could be constructed which created a reasonable prediction. A forecast was created using Holt Linear Exponential smoothing method with coefficients of 0.642 and -0.032 for male data, and 0.747 and -0.097 for female data which was calculated using Microsoft Excel Solver, and these forecasts could be inserted in the Neural Network.

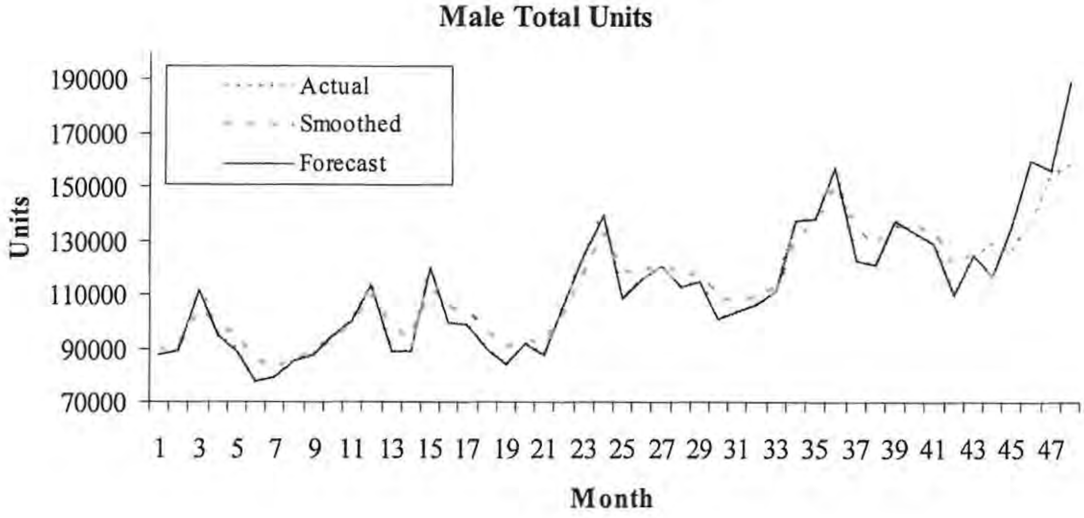


Figure E.3: Male Units Sold

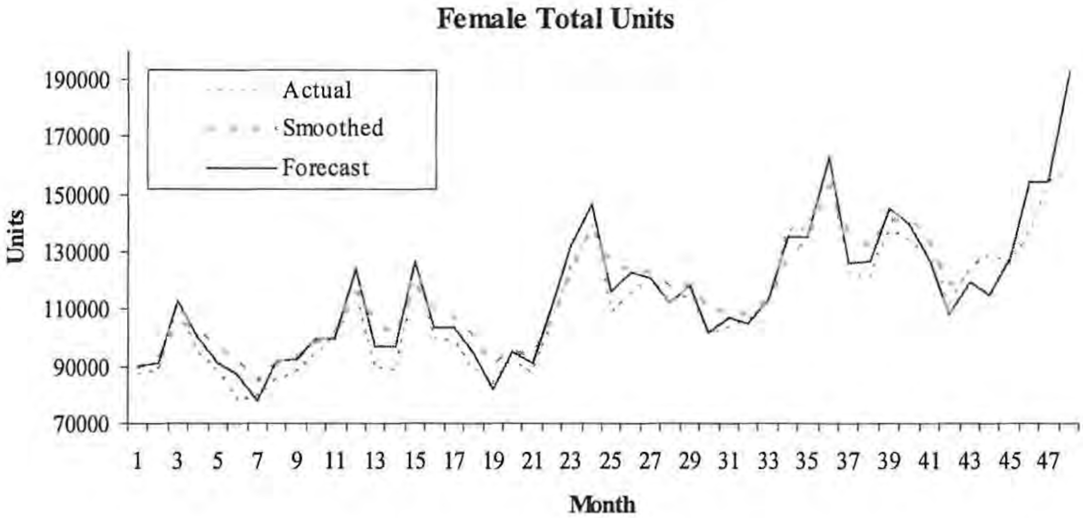


Figure E.4: Female Units Sold

Average Price

The plots of both male and female average price appear to follow a random motion, and cannot be predicted with any certainty. The Neural Network still requires a forecast, so using the previous months average price is recommended as the price does not often have dramatic changes between months.

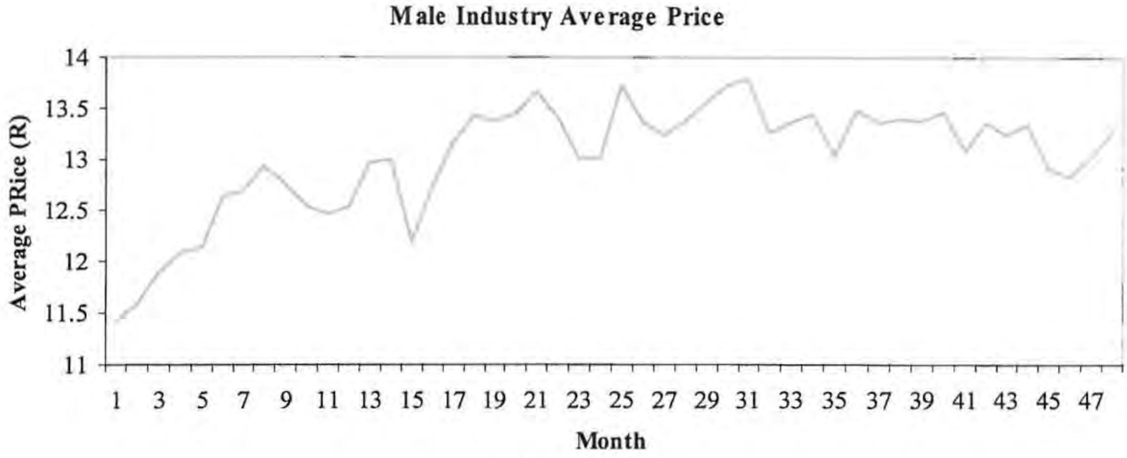


Figure E.5: Male Average Price

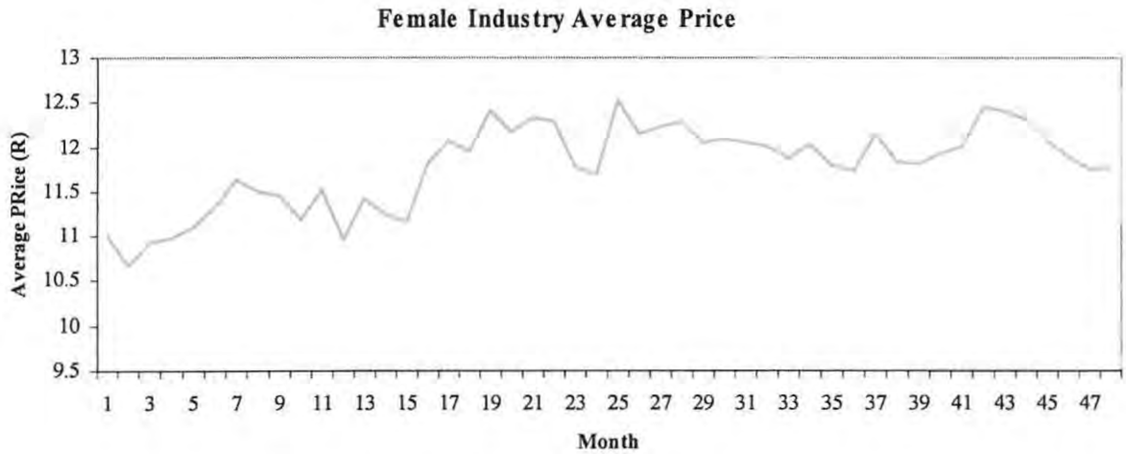


Figure E.5: Female Average Price