Insert here your thesis' task.

Czech Technical University in Prague
Faculty of Information Technology
Department of theoretical computer science

Master's thesis

Predictive Models in Logistics: Comparison of traditional time series techniques with an artificial neural network model approach

Bc. Ruhi Ravichandran

Supervisor: Ing. Pavel Kordik, Ph.D.

3rd May 2015

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my thesis adviser Ing. Pavel Kordik, Ph.D, for his guidance, support and continuous encouragement.

I would also like to thank the anonymous FMCG company and my mentor at that company, for his help in getting the data and providing me with all the necessary information and guidance in order to complete my thesis.

I would like to show my appreciation to the Dean, Vice-Dean for Studies, all my teachers and the Study Department at FIT for helping me out whenever I had any kind of problem with regard to my study plan and making my study period at the Czech Technical University very comfortable and memorable.

Finally, I would like to thank the most important person in my life, my husband Adithya for believing in me and standing by me. This journey would not have been possible without his love and support.

Declaration

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis.

I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular that the Czech Technical University in Prague has the right to conclude a license agreement on the utilization of this thesis as school work under the provisions of Article 60(1) of the Act.

In Prague on 3rd May 2015 ....

Citation of this thesis

Ravichandran, Ruhi. *Predictive Models in Logistics: Comparison of traditional time series techniques with an artificial neural network model approach.* Master's thesis. Czech Technical University in Prague, Faculty of Information Technology, 2015.

Abstrakt

Tato práce zabývá srovnáním tradičních metod analýzy časových řad; konkrétně pak Holt-Wintersovou metodou, exponenciálním vyrovnáváním a autoregresivními intergrovanými modely klouzavých průměrů (ARIMA) s umělou neuronovou sítí za účelem prognózy úrovně zásob více skladových jednotek (SKU - Stock Keeping Unit) na konci dodavatelského řetězce - v maloobchodní prodejně. Komparace je provedena za pomoci různých ukazatelů přesnosti prognóz. Je zde zkoumáno, jaký přístup je vhodné zvolit při prognóze úrovně zásob v prodejně na příkladu společnosti vyrábějící více SKUs. Práce rovněž objasňuje rozličné faktory, které ovlivňují kvalitu zvolených modelů; například události spojené s prodejem jako je reklama, období dovolené, víkendy atp. či četnost prognóz; zda jsou vyhotovovány týdně nebo měsíčně. Modelování a analýza byly provedeny v programovacím jazyku R. Data zde použitá jsou reálná a byla získána od přední společnosti z oblasti rychloobrátkového zboží. Výzkum se proto zaměřuje právě na toto odvětví a nabízí řešení ke zlepšení prognóz poptávky výše uvedené společnosti.

Klíčová slova Statistické modelování, časové řady, ARIMA, Holt Winters, exponenciální vyhlazování, umělé neuronové sítě, předpovědi poptávky..

Abstract

Demand forecasting is a crucial part of managing any supply chain network, since inaccurate forecasting often leads to inventory mismanagement which in-turn amounts to big losses for companies.Though most of the companies have some forecasting techniques in place, it is equally important to know if the forecasting techniques being used are best suited for their requirements.

This thesis provides a comparative study of traditional time series methods namely: *Holt Winters, Exponential smoothing and ARIMA* with an artificial neural network model in order to forecast inventory levels of multiple SKUs at the last mile of the supply chain, which is a retail store. Comparison is performed using various forecasting accuracy measures. The study provides insights to the company that manufactures number of SKUs, as to which forecasting techniques would best suit their need for managing inventory at the store level and why. It also sheds light on the factors that affect the performance of models, for example sequence of events linked to sales like promotions, holiday season, weekends etc. or the granularity at which the forecasting is being done, whether it is weekly or monthly. Modelling and analysis was performed in R programming environment. The data used for this study is real world point of sales data provided by a leading fast moving consumer goods company. It is an industry based research to help the company improve their demand forecasting techniques.

Keywords Statistical modelling, Times series, ARIMA, Holt Winters, Exponential smoothing, Artificial Neural Networks, Demand forecasting.

Contents

List of Figures

List of Tables

Introduction

Motivation and Objectives

In order to gain competitive advantage and maintain strong consumer base in a rapidly evolving business environment driven by changes in shopper habits demand forecasting is very important for any consumer goods company as it helps them to make the right decisions with regard to manufacturing and inventory management.

One of the well reported modern era supply chain disasters was seen in June 2000 when Nike installed a 400 million dollar demand planning software i2, but after a period of 9 months it led to major inventory write offs which in turn led to 100 million dollars in lost sales, depressed its stock price by 20 percent and triggered a flurry of class-action lawsuits. All of this is because they succumbed to the pressure of demand forecasting and thought that throwing a bunch of historical sales numbers into a program would lead to a magic number to emerge from the algorithm[1]. Demand planning is not about just using any forecasting software, as the accuracy of prediction depends on a number of factors, the complexity of the supply chain being studied, collection of the correct data, being prepared to tackle unexpected changes in the market conditions etc., hence the key is to use the correct technique which suits user the best.

In consumer goods and retail, sales forecasting attempts to optimize the trade off between customer demand satisfaction and inventory costs. There has been an emphasis on research in this field because if there is a discrepancy between supply and demand it leads to losses, as overestimation gives way to markdowns, excess inventory and disposal costs, and underestimation leads to stock outs[2]. In both the cases there is a huge loss to the company and the retailer as either they end up selling their products at a lower price to clear excess inventory or they lose loyal customers who are forced to look for alternatives due to stock outs. Companies are constantly looking at ways to improve forecasting, so research on this topic creates a direct relation between the practical implications to the company and also enrich research on the statistical forecasting, which makes accurate forecasting all the more important.

Moreover, forecasting is essential for optimizing revenue and profits[3]. Furthermore, if demand exceeds supply, it appears that forecasts may be bloated and it is unclear how much revenue is being missed. It has been shown that if demand forecasting is improved it may lead to significant monetary savings, improved customer satisfaction, and better channel relationships[4].

Predicting future sales has always been an important task of sales managers and in the past was mainly done by managerial judgment. It is only in the last decade that they have had the support of sophisticated tools and methods to improve their forecasting tactics. As it is becoming more of a science it is firstly important to understand the different methods of forecasting available before deciding on which method to adapt. Different forecasting methods are used because of their different applications and it is of relevance to understand the drivers behind these applications in their specific context[5]. Since each forecasting method has its advantages and disadvantages it is important to look at them separately and select those methods which will prove to be the most useful under specific conditions.

Literature Review

In the retail industry, it is observed that large retailers are more likely to use time-series methods, whereas smaller retailers emphasize on judgemental methods[6].

Over the past decade a number of statistical techniques such as Holt-Winters, Exponential smoothing, ARIMA, Box-Jenkins Models, Regression models have been used to perform demand forecasting, these models are also applied to data containing seasonal trends and patterns.The goal of time series forecasting is to predict the behaviour of complex systems by studying only past records of the same phenomenon[7].

Newer approaches in this field include application of artificial intelligence methods such as artificial neural networks.It has been recommended that ANN be used to investigate how seasonal patterns change over time[8]. ANN algorithms have been found to be useful techniques for demand forecasting due to their ability to be data-driven self adaptive methods. They learn from examples and capture subtle functional relationships among the data even if the underlying relationships are unknown and hard to describe. Thus ANNs are well suited for problems whose solutions require knowledge that is difficult to specify but for which there are enough data or observations[9].

Research has also suggested the use of hybrid models, which are developed using advantages of various models, as a new combined approach to improve forecasting accuracy[10]. Study on time series forecasting using a hybrid AR-

IMA and neural network model, showed how linear ARIMA model and the nonlinear ANN model are used jointly to overcome the drawbacks faced by each of these methods,by fitting the ARIMA model first to the data it was found that the problem of over-fitting which is strongly related to ANN was eased[11].

There have also been comparative study of classical methods, which are based on statistical models and modern heuristic methods (ANN)[12]. As part of the classical methods group they have considered,Holt- Winters, ARIMA, and Regression models. Most of these models are linear and find it difficult to cope with highly erratic real-world sales data[13]. In contrast, it is seen that the modern heuristic methods like artificial intelligence, are able to handle these challenges better.The use of ANN has been gradually increasing over the years in marketing and retailing fields. They have been used in various types of applications, like market response forecasting[14], consumer choice forecasting[15], market segmentation analysis[16] and analysis of buyer seller relationship[17].

But we found some contrasting views in the literature as well, where certain findings have led to mixed response with respect to ANN's potential. It was found that ARIMA model has an equivalent or even slightly better mean absolute percentage error than ANN, but the forecast error for ANN was lower than ARIMA when there were some trends and patterns in the data while analysing a 50-M competition series[18]. Neural networks are supposed to simultaneously detect non linear trend and seasonality in the data[19], but on the other hand it has also been found that ANN did not model seasonal fluctuations in data as well as expected[20]. The instability of different neural network models to directly forecast seasonal time series was also observed during comparative study of airline data, where ANN was pitted against Box-Jenkins method[21].

It was seen that ANN outperformed traditional time series methods when forecasting quarterly and monthly data[22], but in contrast exponential smoothing was found to be better than ANN in forecasting yearly data and almost comparable for forecasting quarterly data[23]. Exponential smoothing method has also proven to be superior in a lot of contexts, like detecting the peak demand of an electric utility company[24], predicting various components of income tax i.e. earnings before taxes, earnings before interest and taxes etc. In the study of income tax prediction ARIMA too proved to an equally strong contender, when compared to Census X-11 and random walk models[25]. Winter's model performs better than Holt's model to forecast aggregate retail sales[26]. ARIMA was found to perform as good as ANN when used in a study of 75 out of 111-series[27] and ARIMA was also found to give very accurate short term forecasts when compared to other econometric models[28].

As it was observed that both ARIMA and exponential smoothing models have proven to be effective time series forecasting techniques and many researchers have used them to compare with the relatively new ANN model.Also the Holt-Winters technique too seems to be a good option for retail sales forecasting.Hence we decided to explore these methods as part of this thesis.

Tasks and Goals

This thesis work is an industry based quantitative research which deals with modelling and analysis of different demand forecasting techniques. The tasks set up for this work can be explained as follows:

- Collection of data from the industry partner and understand their existing forecasting techniques and requirements. This will help in forming the research question which need to be answered.
- Performing relevant literature survey to understand the existing trends in retail forecasting and to select appropriate techniques which will be used in the study.
- As part of the literature survey understand the different forecasting accuracy measures.
- Using the information from the literature survey, set up the necessary theoretical background.
- Preparation of data for the experiments, separation of data into training and test sets.
- Selecting the tool on which the analysis will be done, and performing the modelling of the selected techniques.
- Running all the models using the same training and test sets, computing the error metrics under different conditions and comparing the models on the basis of these metrics.
- Analysis and discussion of the results.
- Writing down the conclusions based on results obtained in the experimental phase and making sure that the research questions are answered.
- Proposing some suggestions to the industry partner, depending on the results obtained from the study.

The goals that I plan to achieve as part of this thesis work are as follows:

- Comparison of traditional time series techniques with ANN and finding which methods are the best suited for their data set I am working with. Try to find the reason why some models perform better than the others and the factors that influence their performance.

- Answering research questions which we formed on the lines of the company's requirements :

    1. *Which methods are most suitable for forecasting sales on a higher granular level?* Since accurate forecasting on a daily or weekly basis would be more helpful to them as most of the time they seem to face difficulties with delivering stock at a short notice from the retailer, and if they are able to forecast the sales at the retailer level in advance they can make direct supply to the point of sale by skipping the mid levels and avoid the time lag and this will help them avoid stock out issues.
    2. *What kind of data do they have to collect and how much data do they need to save, to help them forecast better?* Depending upon the methods that suit their needs the best, they would like to know about the kind of data they need to collect to help them forecast better. Currently they seem to have a lot of data but they do not know if the data being collected is useful and if they are missing out on some really useful data, and how much data do they have to save, because as of now most of their data sets have sales figures for a period of maximum 2 years.

Structure of the thesis

The thesis is divided into three main parts:

- Chapter 1 presents the theoretical background based on the literature survey.
- Chapter 2 gives an insight into the experiments performed on the data set and analysis of the results obtained from the experiments.
- Chapter 3 provides the conclusions drawn from the experiments and the results obtained from them. It also presents the future work that can be done based on this thesis.

Chapter **1**

Theoretical Background

This chapter provides an introduction to fast moving consumer goods and its supply chain management, it also discusses the different forecasting techniques and accuracy measures being used in the study.

1.1 Introduction to FCMG industry

Fast moving consumer goods or FMCG industry also referred to as the Consumer Packaged Goods or CPG industry handles manufacturing, distribution and marketing of CPG. It is a multi-million dollar industry and constitutes a wide range of brands that we use in our everyday life. They are called "fast moving" as they are generally the fastest to go off the shelves in stores, either due to high consumer demand or due to deterioration. These are the items that get used up quickly in our households and need to be replaced at short intervals.

Soaps, detergents and various other cleaning and laundry products, personal care items, pharmaceuticals, food and other dairy products make up a large chunk of the goods in the FMCG sector, but the list doesn't end there. Paper products, consumer electronics, plastic goods, beverages, printing and stationery, alcoholic drinks, tobacco and cigarettes are also some of the most popular FMCG products[29].

FMCG products are generally sold quickly and a relatively low price. Though the profit margins made on these products is relatively low specially for the retailers than for the manufacturers or suppliers, but because they are sold in large quantities the total profit made on these products could be of a considerable amount. FMCG is probably the most classic case of low margin and high volume business[30].

FMCG industry has rapidly grown in the past decades which has led to high competition between FMCG manufacturers. It was also observed that in some countries this industry was among the top performing sectors. In India, the FMCG industry is regarded as the fourth largest sector with total market size of US$ 13.1 billion. Also in New Zealand it is regarded as the largest sector accounting for 5% of Gross Domestic Product (GDP)[29].

Some of the most popular FMCG companies include Procter & Gamble, Unilever, Nestle, Coca-Cola, Reckitt Benckiser, Colgate-Palmolive etc. These FMCG companies are known mostly by the brands they sell, some of the most popular brands sold by them are :

- Procter & Gamble : Ariel, Pampers, Head & Shoulders, Gillette
- Unilever : Dove, Lipton, Hellmann's, Savo
- Nestle : Nescafe, Orion, JOJO, Nestea

1.1.1 Characteristics of FMCG

1. Characteristics from a customers perspective
   - Frequent purchase : Goods that are frequently purchased by the consumer, example salt, sugar, bread, eggs. These items are never stocked beyond the required quantity as they easily available and are inexpensive.
   - Low involvement : FMCG's are low involvement goods as in most of the cases if the customers choice of brand is not available, there is a high chance that he or she will take the alternate options offered in its place due to the availability of a huge variety of options, for example in the case of ketchup or salt. But in this case there can be exceptions where a customer is comfortable using a certain brand, then he or she would not compromise and go to a different store to buy the brand of their choice, this is mainly observed in case of personal hygiene products.
   - Low price : These goods are usually low priced, though sometimes the customers finds certain brands prices higher when compared to the others, but in case most of the frequently bought items like potato wafers, biscuits, etc the price differences are very low and so are the costs. This is another reason why the customer does not bother with choice and shows low involvement.
2. Characteristics from a marketers perspective
   - High Volumes : FMCG industry is all about high volumes. For example a family of 4 might use 2-3 cakes of soap per month and so if this is multiplied for the all the families available of the same size across the city, the number of soap cakes required is very huge. The same goes for toilet paper, tooth pastes etc. Hence the manufacturer must be prepared to handle large volumes of sales.

- Low margins : Due to huge competition in the sector, these products are usually sold at costs which are very close to their production costs. But to high volumes of sale the turnover is generally good.
- Extensive distribution networks : Due to low level of involvement at the customers end, it is seen that customers in most cases are not loyal to any particular brand and hence they buy whatever is available or something that saw in an advertisement which catches their imagination or in case of impulse buying they end up buying whatever brand they see. Hence manufacturers must make sure that their brands are well distributed.
- High stock turnover : Due to large volumes being sold and the high frequency of sales, retailers are willing to trade in these products as they help turn his capital almost daily[31].

1.2 FMCG supply chain

Supply chain refers to a sequence of events which take place in order to transform the raw material into a finished product and make it accessible to the end user. It consists of a network of entities who are either directly or directly involved in this process. It comprises of raw material suppliers, manufacturers, warehouses, distribution centers and retailers. It not only consists of organizations and people, but also consists of activities which take place throughout chain and information which is passed from one phase to the other so that the supply chain can be improved. Supply chain management is an important process for companies these days, as no matter how good their product is, if their supply chain is not optimum they can lose out to competition in terms of popularity as well as revenue. Hence companies are constantly striving to improve their supply chain, as it not only improves their sales but also helps them lower their costs.

Supply chain can sometimes become complex as well depending on how many entities are involved in the process. And they can never be generalized, even if it for the same industry or for that matter the same brand of products. Lets consider the example of FMCG industry, each unit of product in retail industry is called as a stock keeping unit (SKU), in simple terms we can say that an ariel detergent 30 wash blue packet is one SKU, ariel detergent 30 wash box another SKU, ariel liquid 10 wash bottle a different SKU and ariel detergent tablets 30 wash a different one and so on. This shows that SKU's can differ on various factors colour of packets, state of the product (solid/liquid), weight etc. While calculating the inventory, the retailer or distributor always counts the quantities in terms of SKU's. In FMCG industry the SKU's are classified into 3 main categories :

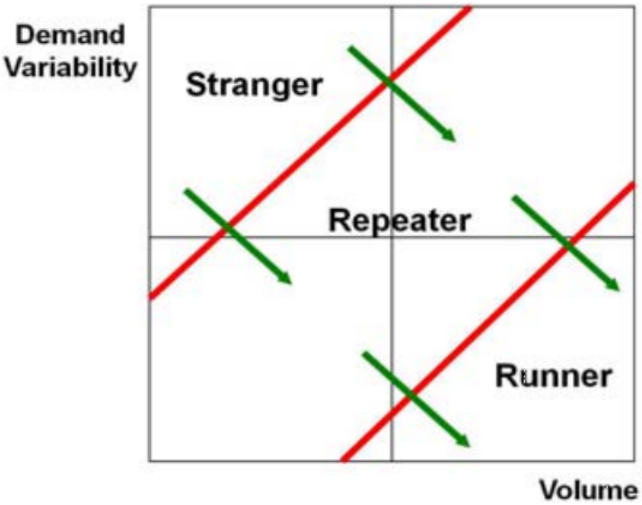

Figure 1.1: SKU classification

1. Runners : These are SKU's which are purchased in high volumes and the variability of the demand is very low.

2. Strangers : These SKU's are purchased in very low volumes and the demand is highly variable.

3. Repeaters : SKU's which fall between runners and strangers, these have intermediate level of demand and variability[32].

So generally if the distribution center is shipping products to the retailers for 10 days, it is seen that runner products are shipped 80% of the days ie on 8 out of the 10 days, the stranger products are shipped 5% of the days and the repeaters fall in between runners and strangers[33]. Due to such large variations in demand it is almost impossible to generalize supply chain. However we try to explain how the supply chain in a modern selling channel works.

We distinguish two different flows in the supply chain, a material and an information flow.The material flow streams from the upper node in the supply chain (raw material supplier) towards the lowest node (the end consumer). The information flow follows the opposite way. The information flow starts and the material flow finishes when the end consumer buys a product in the retailer's store.

Since retailers need product availability in the shelves, product sales will trigger the store to place an order at its distribution centre (DC). The DC accordingly receives orders of multiple stores and delivers the products to its stores (material flow). Subsequently, deliveries from the DC towards the

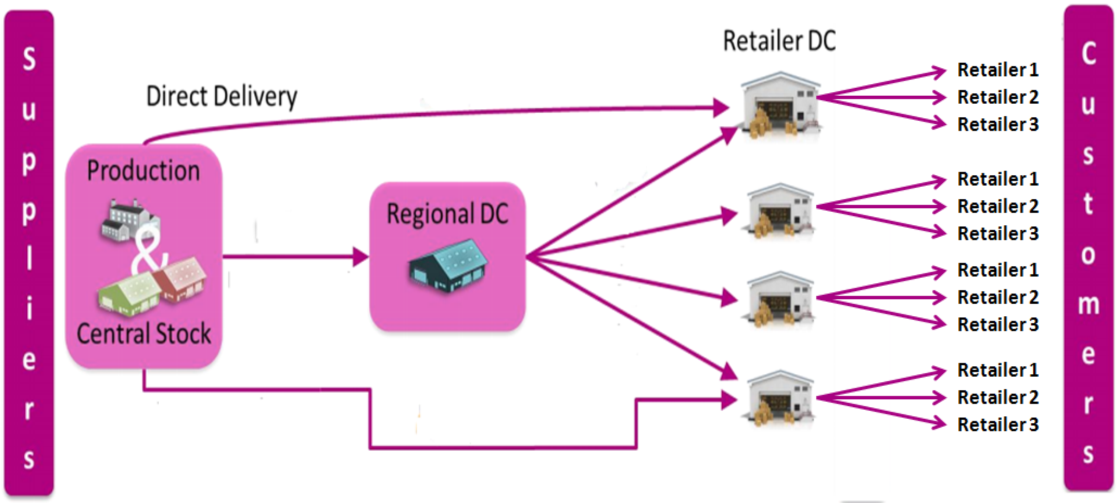

Figure 1.2: Product supply chain

stores trigger the DC to order new products at the manufacturers regional DC (information flow). The manufacturers regional DC similarly collects orders from different retailer DC's and delivers the products to them. In turn it places an order at its plant. Each plant mostly produces several brands of one category. It depends on the product category if production is forecasted or based on real orders or a combination of these two. After production the finished goods will be held in warehouses at the plant, before being dispatched to their regional centers.The finished products are generally held in inventory either at the plant warehouse or at the regional DC till they are delivered to the retailer. The company holds the raw materials on stock till they are ready for production. Manufacturers want to ensure continuous production at the plant since a production pause leads to problems throughout the downstream supply chain and can lead to missed sales. Therefore availability of materials needed for production is really important. Deliveries from raw material suppliers to manufacturers are often based on forecasts rather than on orders. In some cases the finished products are directly shipped from the plant warehouse to either the retailer DC or in some cases to the retailer stores. It depends on the agreement between the retailer and the manufacturer, sometime if the order is urgent or there are bulk orders or the ordered items are of very high value then this direct delivery method is adopted.

1.2.1 Challenges faced by FMCG supply chain

The key issue in FMCG supply chain is to maintain the level of logistics that satisfies the requirements of order fulfillment, both time and cost aspects concurrently[34]. Accurate customer demand forecasting is very important because it can help to :

- Ensure product availability : Making sure that the products are available on the shelves of the store ensures a happy and satisfied customer, which keeps him loyal to the product or brand and avoids missed sales due to product stock out.
- Avoid over production and over stocking : Ensuring adequate stocks can help in avoiding marking down of the price of the product in order to clear excess stock, and also reduces inventory costs.
- Reduce the lead time : Lead time refers to the time between placement of the order and the delivery. With vendor managed inventory (VMI), where the manufacturer is managing the inventory at the retail store level and has access to all the point of sale (POS) data, if the manufacturer forecasts the demand accurately then he can ensure stock availability and deliver it to the retailer at the earliest.

But sometimes it becomes very difficult to have accurate forecasting due to the following reasons:

- Inventory inaccuracy: Incorrect inventory records can give a wrong estimation of the items sold, which is a big hindrance to accurate forecasting. There can be a lot of reasons which lead to inaccurate inventory, some of which are stock loss due to theft, transaction errors which can either happen due to errors in shipment records or errors at the cash registers while scanning the products, inaccessible inventory which occur as a result of not being able to find certain articles as their location is unknowing due to incorrect placements by staff or customers, incorrect product identification due to placement or wrong labels.
- Use velocity: As the use velocity can differ from one buyer to another, it becomes difficult to predict accurately when buyers would return to restock the product. For instance, buyer "A" who has a family of 4 might replenish some product after 2 months but the same product will be replenished by buyer "B" who has a family of only 2 members in maybe 5 months.
- Increase or decrease in price: If the manufacturer decided to increase the price of a product, this may lead to some section of buyers to shift to other brands of the same product which are comparatively cheaper as excess costs might upset their budgets, or if the retailer decides to mark down the prices on products due to promotions this lead to customers buying more than their requirement and they end up storing the products. These factors disrupt the customers regular buying cycles, and hence can cause troubles while forecasting.

- Competitor activity: If a competitor launches a new product sometimes a large chunk of buyers might move to the new product due to many reasons, maybe they find the new product more attractive or it works better for the needs. Heavy promotions on the competitor side is also another reason for customers to shift base. These factors can also affect forecasting as it is difficult to know when competitors are going to launch new products or how many loyal consumers would shift base.

There can be many other reasons which might change consumer buying trends, and any such change is sure to have an effect on demand forecasting.

1.3 Forecasting techniques

Forecasting techniques are broadly classified as qualitative and quantitative methods, the classification is mainly dependent on the kind of data that is available.

- Qualitative forecasting: These methods are adopted when either no data is available or the data available is not relevant to the forecasts. They are done based on opinions and judgements. Some examples are delphi technique, forecasting by analogy, consumer surveys etc.

- Quantitative forecasting: These methods can be applied when numerical information about the past is available and it can be assumed that some past patterns will continue in the future. Most of these methods use either time series data which are collected over regular time intervals or cross-sectional data which are collected at a single point in time[35]. Quantitative approaches can be further classified as:

    1. Time series methods, which use historical data to forecast the future demand. We will discuss these methods in detail in the next section.
    2. Causal methods, which explore the relationship between factors that affect the demand, regression methods are one such example.

In addition to these techniques there has also been a rise in usage of new methods, which are based on machine learning, they include artificial intelligence methods like artificial neural networks, support vector machines etc.

As we will be using time series techniques and artificial neural network method in the implementation, these models are explained in detail in the next section.

1.3.1 Time series forecasting

Time series forecasting is a method that uses historical data to predict the future behaviour of any sequence of observations. These are most useful while forecasting something that is changing over time, like stock prices, sales figures etc[35]. In this implementation we will be considering time series that are observed at regular intervals of time i.e daily, weekly, monthly and yearly.

There are 3 types of time series patterns :

- Trend: A trend exists when there is long term increase or decrease in the data. It does not have to be linear.
- Seasonal: A seasonal pattern exists when a series is influenced by seasonal factors like day of the week, month of the year etc. Seasonality is always of fixed and known period.
- Cyclic: A cyclic pattern exists when data exhibit rises and falls that are not of fixed period. The duration of these fluctuations is usually of at least 2 years[13].

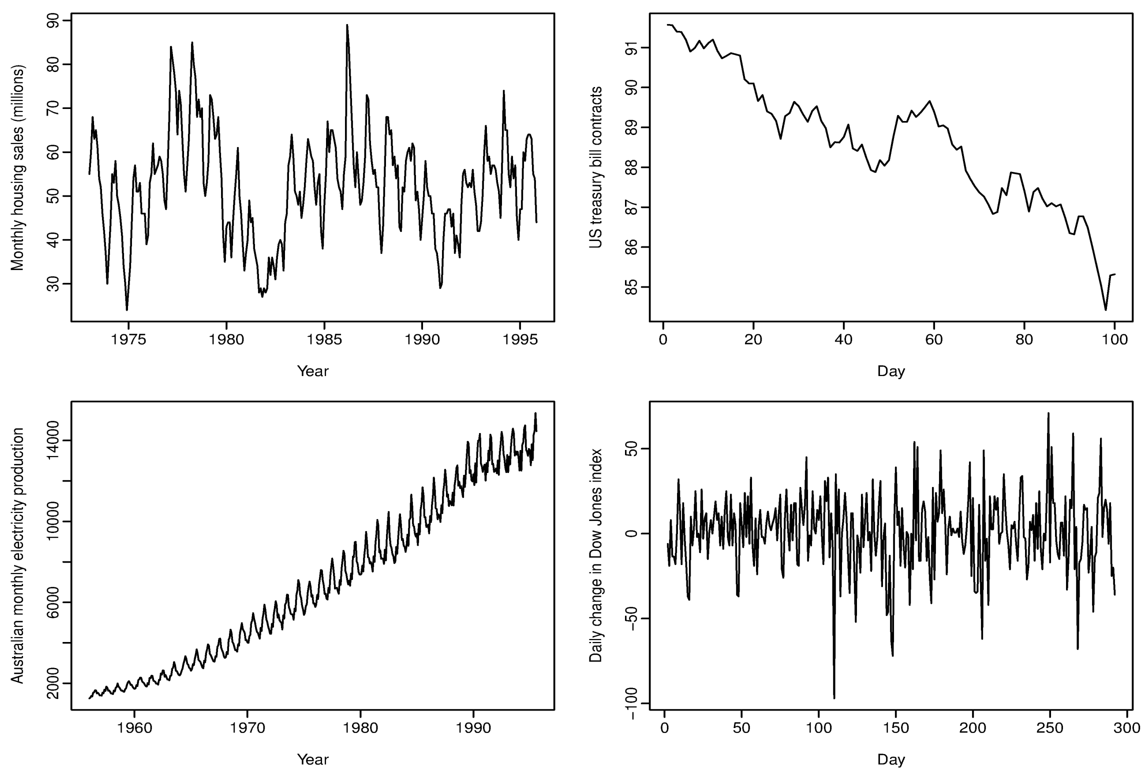

Figure 1.3: Time series showing different types of time series patterns

- The monthly housing sales (top left) shows strong seasonality within each year, as well as some strong cyclic behaviour with period about 6–10 years. There is no apparent trend in the data over this period.

- The US treasury bill contracts (top right) show results from the Chicago market for 100 consecutive trading days in 1981. Here there is no seasonality, but an obvious downward trend. Possibly, if we had a much longer series, we would see that this downward trend is actually part of a long cycle, but when viewed over only 100 days it appears to be a trend.

- The Australian monthly electricity production (bottom left) shows a strong increasing trend, with strong seasonality. There is no evidence of any cyclic behaviour here.

- The daily change in the Dow Jones index (bottom right) has no trend, seasonality or cyclic behaviour. There are random fluctuations which do not appear to be very predictable, and no strong patterns that would help with developing a forecasting model[35].

The following subsections will give a brief explanation about some of the time series techniques used in this implementation.

1.3.1.1 ARIMA

ARIMA (Autoregressive Integrated Moving Average) models also known as Box-Jenkins models are a class of linear models which can represent both stationary as well as non stationary time series. A stationary time series is one whose properties do not depend on the time at which the series is observed, hence they will not have any predictable pattern in the long run. Cyclic time series are considered stationary since the cycles are not of any fixed length here. Whereas time series with trends and seasonality are non stationary since these factors decide the value of time series at different times. A time series can also be made stationary by a process called as Differencing, which computes the differences between consecutive observations. Differencing helps stabilize the mean of a time series by removing changes in the levels of a time series and hence eliminating trends and seasonality. We can also perform seasonal differencing by finding out the difference between an observation and its corresponding value in the previous year[35]. It can be explained by the equation given below:

$$y_t^{'} = y_t - y_{t-m} \tag{1.1}$$

where m is the number of seasons

These are also called "lag m differences" as the observations are subtracted after a lag of m periods.

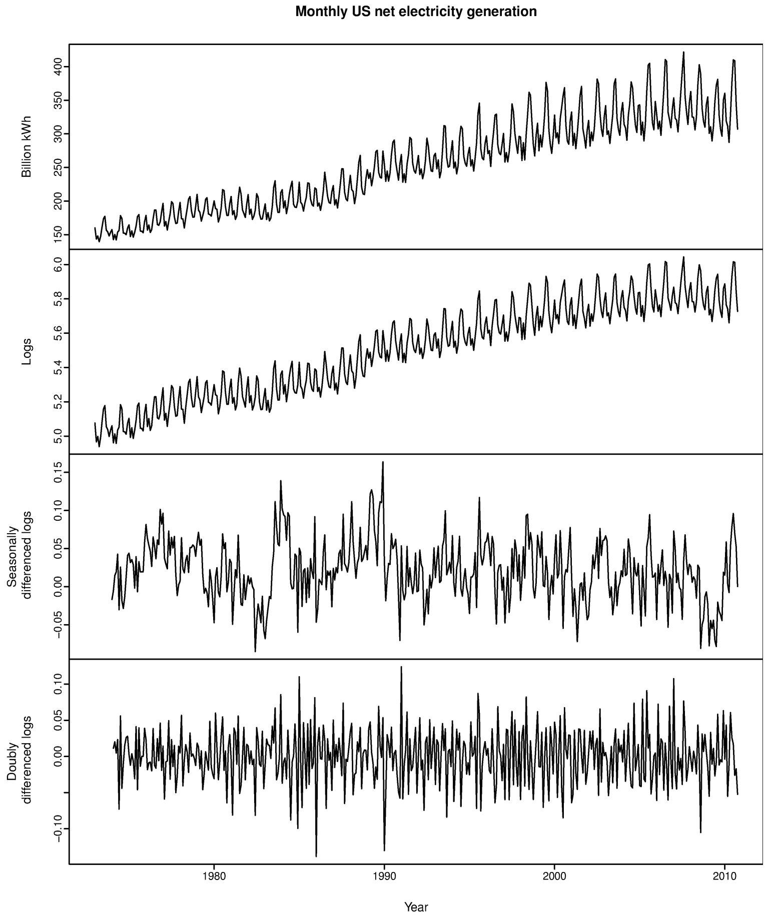

Figure 1.4: Example showing effects of differencing

Figure 1.4 shows the US net electricity generation (billion kWh) in the top panel and the changes in data after applying differencing in the other panels[35].

ARIMA is basically a combination of autoregressive (AR) and moving averages (MA) techniques of times series modelling. These models are flexible and can handle a wide range of different time series patterns.

In an autoregressive model the chosen variable is forecasted using a linear combination of past values of the variable. Autoregression refers to the regression of the variable against itself. Thus an autoregressive model can be of order p can be written as follows :

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + e_t \tag{1.2}$$

where c is a constant and e_t is noise[13].

Figure 1.5 shows examples of autoregressive models with different data.

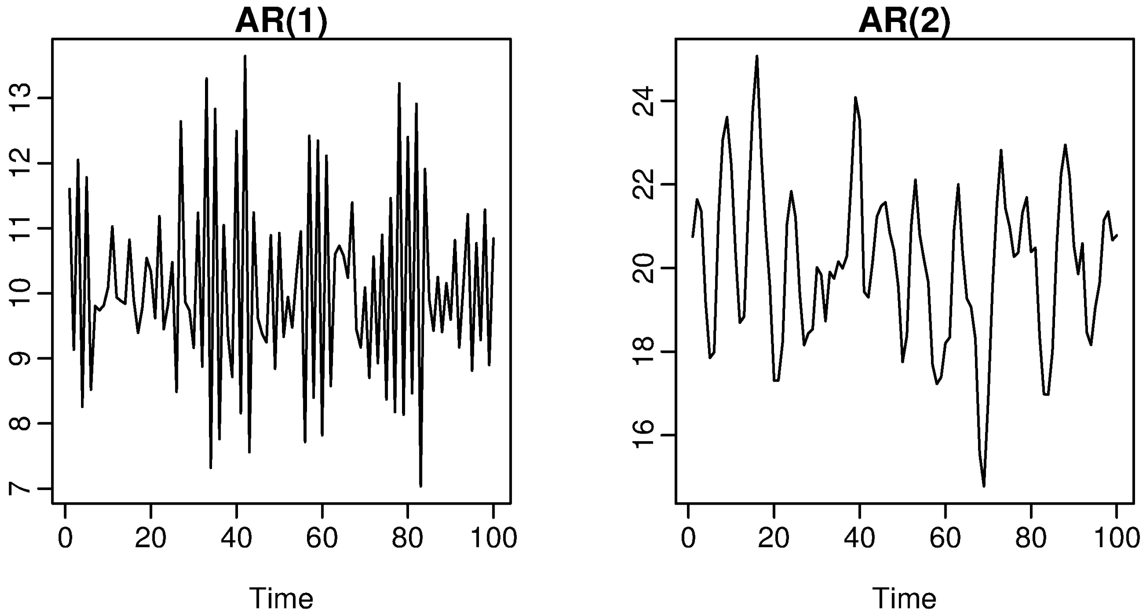

Figure 1.5: AR models with different data

AR(1) with y_t=18-0.8y_{t-1}+e_t

AR(2) with y_t=8+1.3y_{t-1}-0.7y_{t-2}+e_t

In both cases, e_t is normally distributed noise with mean zero and variance one.

It is observed that changing the parameters ϕ_1,...,ϕ_p results in different time series patterns. The variance of the error term e_t will only change the scale of the series, not the patterns[35].

On the other hand moving average model uses past forecast errors in a regression like model. A moving average model of order q can thus be written as :

$$y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + ... + \theta_q e_{t-q} \tag{1.3}$$

where e_t is noise. Each value of y_t can be considered as a weighted moving average of past few forecast errors[13].

Figure 1.6 shows examples of moving averages models with different data.

MA(1) with y_t=20+e_t-0.8e_{t-1}

MA(2) with y_t=e_t-e_{t-1}+0.8e_{t-2}

In both cases, e_t is normally distributed noise with mean zero and variance one.

Similar to autoregression, it is observed that changing the parameters θ_1,...,θ_q results in different time series patterns. The variance of the error term e_t will only change the scale of the series, not the patterns[35].

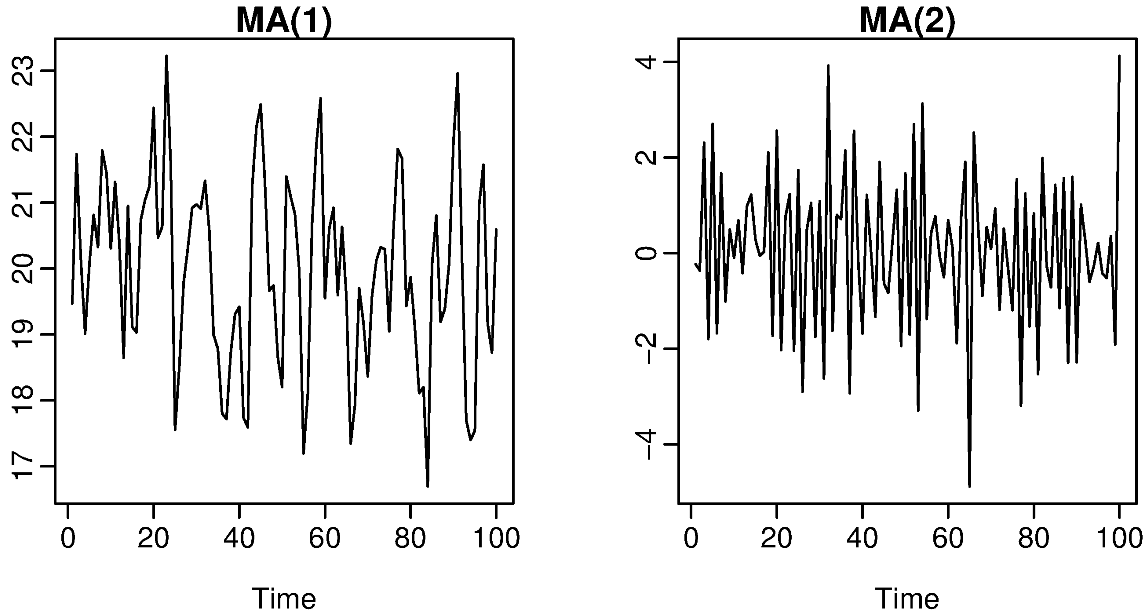

Figure 1.6: MA models with different data

Combining differencing with autoregression and moving averages results in a non seasonal ARIMA model. The full model xan be written as follows :

$$y_t^{'} = c + \phi_1 y_{t-1}^{'} + ... + \phi_p y_{t-p}^{'} + \theta_1 e_{t-1} + ... + \theta_q e_{t-q} + e_t \tag{1.4}$$

where $y_t^{'}$ the differenced series (it may have been differenced more than once). The "predictors" on the right hand side include both lagged values of y_t and lagged errors. This is called as ARIMA(p,d,q) model, where

p= order of the autoregressive part
d= degree of first differencing involved
q= order of the moving average part[13]

Some of the models discussed earlier are special cases of ARIMA model.
Autoregression ARIMA(p,0,0)
Moving average ARIMA(0,0,q)
White noise ARIMA(0,0,0)

ARIMA models are also capable of modelling a wide range of seasonal data. A seasonal data ARIMA is formed by including an additional seasonal term in the ARIMA model seen above. It can be written as :
$ARIMA(p, d, q)(P, D, Q)_m$
where m is the number periods per season[13]

1.3.1.2 Exponential Smoothing

Exponential smoothing, also known as simple exponential smoothing is a very old forecasting techniques proposed way back in the 1950's. Forecasts produced using this method are weighted averages of past observations, with weights decreasing exponentially as the observations get older, it means that the more recent the observations is, the higher will be the weight associated with it[13].

It is explained by the equation given below:

$$y^{\wedge}_{T+h|T} = \alpha y_T + \alpha(1-\alpha)y_{T-1} + \alpha(1-\alpha)^2 y_{T-2} + ... \tag{1.5}$$

where $0 \leq \alpha \leq 1$ is the smoothing parameter. The one step ahead forecast for time $T+1$ is a weighted average of all the observations in the series $y_1, \ldots, y_T$. The rate at which the weights decrease is controlled by the parameter α[35].

Table 1.1 shows how the weights are varying for different values of α when forecasting using simple exponential smoothing. The sum of the weights even for a small α will be approximately one for any reasonable sample size[35].

Table 1.1: Exponential smoothing parameter behaviour

Observation	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$
y_T	*0.2*	*0.4*	*0.6*	*0.8*
y_{T-1}	*0.16*	*0.24*	*0.24*	*0.16*
y_{T-2}	*0.128*	*0.144*	*0.096*	*0.032*
y_{T-3}	*0.1024*	*0.0864*	*0.0384*	*0.0064*

It is observed that for any α between 0 and 1, the weights for different observation decrease with time.

Weighted average form is one of the most popular forms of simple exponential smoothing.

The forecast at time $t+1$ is equal to a weighted average between the most recent observation y_t and the most recent forecast $y^{\wedge}_{t|t-1}$.

For t=1,....T where $0 \leq \alpha \leq 1$ is the smoothing parameter. Let us denote the first forecast y_1 as l_0. Then

$$y^{\wedge}_{2|1} = \alpha y_1 + (1-\alpha)l_0 \tag{1.6}$$

$$y^{\wedge}_{3|2} = \alpha y_2 + (1-\alpha)y^{\wedge}_{2|1} \tag{1.7}$$

$$y^{\wedge}_{4|3} = \alpha y_3 + (1-\alpha)y^{\wedge}_{3|2} \tag{1.8}$$

and so on, thus this lead to

$$y^{\wedge}_{T+1|T} = \alpha y_T + (1-\alpha)y^{\wedge}_{T|T-1} \tag{1.9}$$

Substituting each if these equations in to the following, we get

$$\hat{y}_{3|2} = \alpha y_2 + \alpha(1-\alpha)y_1 + (1-\alpha)^2 l_0 \tag{1.10}$$

$$\hat{y}_{4|3} = \alpha y_3 + \alpha(1-\alpha)y_2 + \alpha(1-\alpha)^2 y_1 + (1-\alpha)^3 l_0 \tag{1.11}$$

and so on, thus we get

$$\hat{y}_{T+1|T} = \sum_{j=0}^{T-1} \alpha(1-\alpha)^j y_{T-j} + (1-\alpha)^T l_0 \tag{1.12}$$

Hence the weighted average form leads to the same forecast equation (1.5)[35]

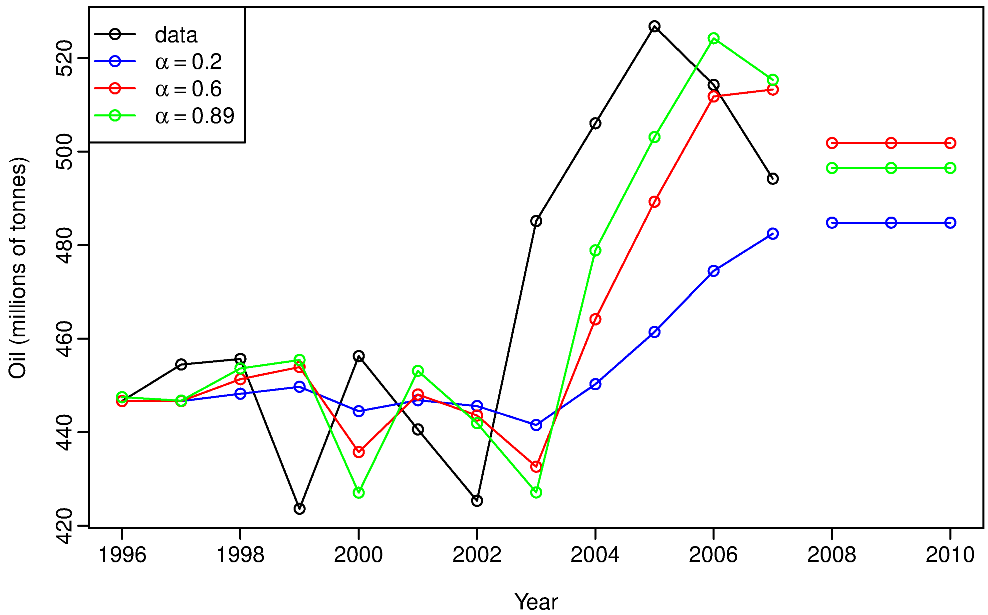

Figure 1.7: Example of simple exponential smoothing

Figure 1.7 shows an example of simple exponential smoothing applied to oil production in Saudi Arabia. The plot for data over the period 1996 to 2007 represented by the black line shows changing behaviour but no obvious trends[35]. It is observed that as the value of the smoothing parameter is increasing the trends are becoming more and more sparse in the plot.

1.3.1.3 Holt-Winters

Holt and Winters is another form of exponential smoothing technique that encorporates seasonality by adding a smoothing term to the trend. This method constitutes the forecasting equation and three smoothing equations, for the level denoted by l_t, for trend denoted by b_t and for seasonal component denoted by s_t, with smoothing parameters α, β, γ[13].

There are 2 variations of this method that differ in the nature of the seasonal component:

1. Holt-Winters additive method: This method is preferred when the the seasonal variations are roughly constant throughout the series. The seasonal component is expressed in absolute terms in the scale of the observed series, and in the level equation the series is seasonally adjusted by subtracting the seasonal component. Within each year the seasonal component will add up to approximately zero. The component form of additive model is given by:

$$y^{\wedge}_{t+h|t} = l_t + hb_t + s_{t-m+h_m^+}$$
$$l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1})$$
$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$$
$$s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}$$

   where h_m^+ ensures that the estimates of the seasonal indices used for forecasting come from the final year of the sample. The level equation shows a weighted average between the seasonally adjusted observation $(y_t - s_{t-m})$ and the non-seasonal forecast $(l_{t-1} + b_{t-1})$ for time t. The trend equation is identical to Holt's linear method. The seasonal equation shows a weighted average between the current seasonal index, $(y_t - l_{t-1} - b_{t-1})$, and the seasonal index of the same season last year[35].

2. Holt-Winters multiplicative method: This method is preferred when the seasonal variations are changing proportional to the level of the series. the seasonal component is expressed in relative terms (percentages) and the series is seasonally adjusted by dividing through by the seasonal component. Within each year, the seasonal component will sum up to approximately m which is the period of seasonality. The component form of additive model is given by:

$$y^{\wedge}_{t+h|t} = (l_t + hb_t)s_{t-m+h_m^+}$$
$$l_t = \alpha\frac{y_t}{s_{t-m}} + (1-\alpha)(l_{t-1} + b_{t-1})$$
$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$$
$$s_t = \gamma\frac{y_t}{(l_{t-1} + b_{t-1})} + (1-\gamma)s_{t-m}[35]$$

Figure 1.8 shows Holt-Winters method with both additive and multiplicative seasonality for forecasting tourist visits at night in Australia by international arrivals.The data shows seasonality with peaks in March each year as this

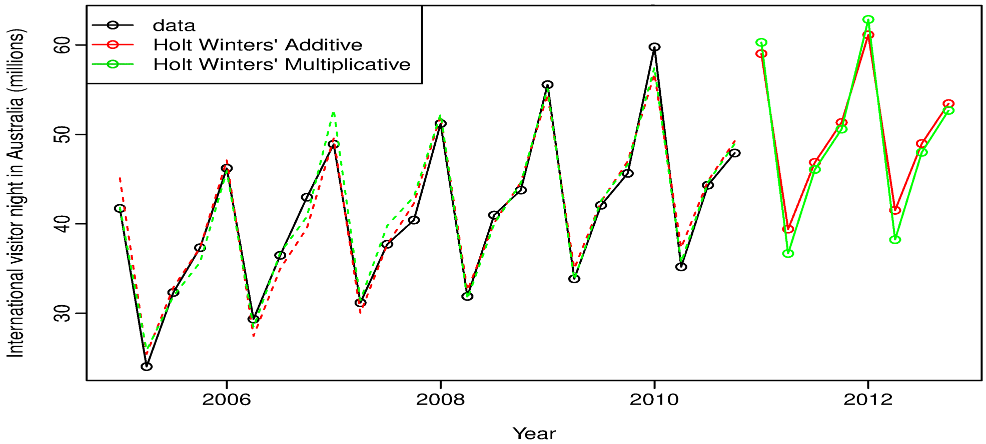

Figure 1.8: Example of Holt-Winters method with both additive and multiplicative seasonality

corresponds to summer season in Australia. Also it can be seen that the multiplicative model fits the data better, since the plot shows seasonal variation in the data increases as the level of series increases[35].

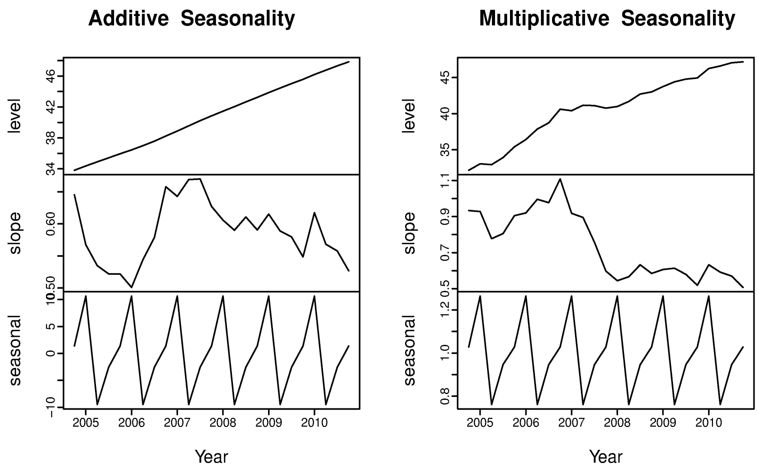

Figure 1.9: Estimated components for Holt-Winters method with both additive and multiplicative seasonality

Figure 1.9 shows the level, slope and seasonal components calculated for the Australian tourist data for both additive and multiplicative seasonality[35].

1.3.2 Artificial Neural Networks

Artificial neural networks are forecasting methods that are based on simple mathematical models of the brain. There are many distinguishing features of ANN which make them attractive forecasting methods, some of these features are listed below :

- They are data driven self adaptive methods.They can learn from examples and capture subtle relationships among data even if the underlying relationships are unknown or hard to describe.
- They can generalize, i.e after learning from sample data presented to them they can correctly infer the unseen parts of a series even if the sample data contains noisy information.
- They are universal function approximators. It has been shown that a neural network can approximate any continuous function to any desired accuracy.
- Unlike traditional approaches they are suitable for forecasting complex nonlinear relationships between the response variable and its predictors[12].

1.3.2.1 An overview of ANN

A neural network can be thought of as a network of "neurons" organised in layers. The predictors (or inputs) form the bottom layer, and the forecasts (or outputs) form the top layer. There may be intermediate layers containing "hidden neurons". The very simplest networks contain no hidden layers and are equivalent to linear regression.

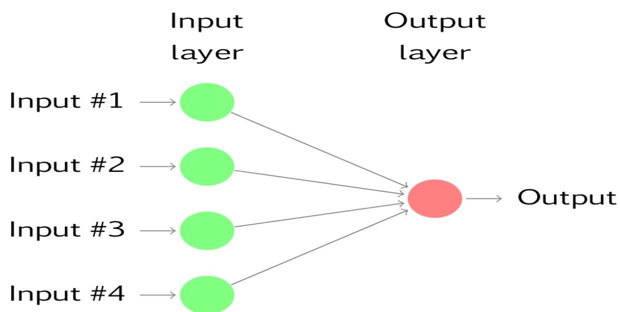

Figure 1.10: Simple neural network

Figure 1.10 shows the neural network version of a linear regression with four predictors. The coefficients attached to these predictors are called "weights". The forecasts are obtained by a linear combination of the inputs. The weights are selected in the neural network framework using a "learning algorithm" that minimises a “cost function” such as MSE. Of course, in this simple example, we can use linear regression which is a much more efficient method for training the model[35].

Once we add an intermediate layer with hidden neurons, the neural network becomes non-linear. A simple example is shown in the figure 11.

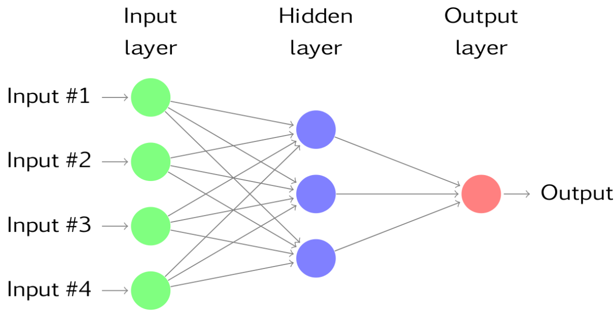

Figure 1.11: Simple multilayer feed-forward network

Figure 1.11 shows a 4-3-1 multilayer feedforward network which has 4 input neurons, 3 hidden neurons and one output neuron.The outputs of nodes in one layer are inputs to the next layer. Each node receives an input signal from other nodes processes it locally through an activation or transfer function and produces a transformed output signal to other nodes[12]. For example, the inputs into hidden neuron j in Figure 1.11 are linearly combined to give

$$z_j = b_j + \sum_{j=1}^{4} w_{i,j} x_i \tag{1.13}$$

where
$w_{i,j}$ refers to the weights associated with the neurons
x_i refers to the input signals
In the hidden layer, this is then modified using a nonlinear function such as a sigmoid,to give the input for the next layer.

$$s(z) = \frac{1}{1 + e^{-z}} \tag{1.14}$$

1.3.2.2 Training of ANN

Before an ANN can perform any task, it must be trained to do so. Training is the process of determining the arc weights which are the key elements of an ANN. The knowledge learned by a network is stored in the arcs and nodes in the form of weights and node biases. It is through the linking arcs that an ANN can carry out complex nonlinear mappings from its input nodes to its output nodes. An multi layer perceptron (MLP) training is a supervised one in which the desired response of the network (target value) for each input pattern (example) is always available. The training input data is in the form of vectors of input variables or training patterns. Corresponding to each element in an input vector is an input node in the network input layer. Hence the number of input nodes is equal to the dimension of input vectors. For a time series forecasting problem, however, the appropriate number of input nodes is not easy to determine. Whatever the dimension, the input vector for a time series forecasting problem will be almost always composed of a moving window of fixed length along the series. The total available data is usually divided into a training set and a test set. The training set is used for estimating the arc weights while the test set is used for measuring the generalization ability of the network.

The training process is usually as follows. First, examples of the training set are entered into the input nodes. The activation values of the input nodes are weighted and accumulated at each node in the first hidden layer. The total is then transformed by an activation function into the node's activation value. It in turn becomes an input into the nodes in the next layer, until eventually the output activation values are found. The training algorithm is used to find the weights that minimize some overall error measure such as the sum of squared errors (SSE) or mean squared errors (MSE). Hence the network training is actually an unconstrained nonlinear minimization problem.

The most popularly used training method is the backpropagation algorithm which is essentially a gradient steepest descent method. For the gradient descent algorithm, a step size which is called the learning rate must be specified. The learning rate is crucial for backpropagation learning algorithm since it determines the magnitude of weight changes. It is well known that the steepest descent suffers the problems of slow convergence, inefficiency, and lack of robustness. Furthermore it can be very sensitive to the choice of the learning rate. Smaller learning rates tend to slow the learning process while larger learning rates may cause network oscillation in the weight space[12].

For a weight connecting a node in layer k to a node in layer j, the change in weight in case of a backpropagation algorithm is given by

$$\Delta w_{kj}(n) = \alpha \delta_j y_k + \eta \Delta w_{kj}(n-1) \tag{1.15}$$

where:

- α is the learning rate, a real value on the interval (0,1]
- y_k is the activation of the node in layer k, i.e. the activation of the presynaptic node, the one upstream of the weight
- n and $n-1$ refer to the iteration through the loop
- η is the momentum, a real value on the interval [0,1)
- δ_j is the "error term" associated with the node after the weight, i.e. the postsynaptic node.

If j is the output layer, then

$$\delta_j = (t - y_j)y_j(1 - y_j) \tag{1.16}$$

If j is the output layer, then

$$\delta_j = (\sum_{i \in I_j} \delta_i w_{ji})y_j(1 - y_j) \tag{1.17}$$

It can be seen in Eq(1.17) that calculation of the error term for a node in a hidden layer requires the error term from nodes in the subsequent (i.e. downstream) layer, and so on until the output layer error terms are calculated using Eq(1.16). Thus, computation of the error terms must proceed backwards through the network, beginning with the output layer and terminating with the first hidden layer. It is this backwards propagation of error terms after which the algorithm is named.

1.4 Performance Metrics

Performance metrics are methods which are used to check the accuracy of forecasted values relative to the original values. Some of the metrics are explained below :

Let y_i denote the i^{th} observation and $y_i^{\wedge}$ denote a forecast of y_i.

The forecast error is given by

$$e_i{=}y_i{-}y_i^{\wedge}$$

1. Scale dependent errors

   There are some commonly used accuracy measures whose scale depends on the scale of the data. These are useful when comparing different methods applied to the same set of data, but should not be used, for example, when comparing across data sets that have different scales[36].

   The most commonly used scale-dependent measures are based on the absolute error or squared errors:

   $$\text{Mean absolute error (MAE)} = \text{mean } (|e_i|)$$
   $$\text{Root mean squared error (RMSE)} = \sqrt{mean(e_i)^2}$$

   When comparing forecast methods on a single data set, the MAE is popular as it is easy to understand and compute[35].

2. Measures based on percentage errors

   The percentage error is given by $p_i = 100e_i/y_i$. Percentage errors have the advantage of being scale independent, and so are frequently used to compare forecast performance across different data sets.

   The most commonly used measure based on percentage errors is :

$$\text{Mean absolute percentage error (MAPE)} = mean(|p_i|)$$

   Measures based on percentage errors have the disadvantage of being infinite or undefined if $y_i = 0$ for any i in the period of interest, and having extreme values when any y_i is close to zero. Another problem with percentage errors that is often overlooked is that they assume a meaningful zero. For example, a percentage error makes no sense when measuring the accuracy of temperature forecasts on the Fahrenheit or Celsius scales. They also have the disadvantage that they put a heavier penalty on negative errors than on positive errors[36].

3. Scaled errors

   Scaled errors were proposed as an alternative to using percentage errors when comparing forecast accuracy across series on different scales. They proposed scaling the errors based on the training MAE from a simple forecast method.

   For a non-seasonal time series, a useful way to define a scaled error uses naive forecasts:

$$q_j = \frac{e_j}{(\frac{1}{T-1})\sum_T^{t=2}|y_t - y_{t-1}|}$$

   Because the numerator and denominator both involve values on the scale of the original data, q_j is independent of the scale of the data. A scaled error is less than one if it arises from a better forecast than the average naive forecast computed on the training data. Conversely, it is greater than one if the forecast is worse than the average naive forecast computed on the training data. For seasonal time series, a scaled error can be defined using seasonal naive forecasts:

$$q_j = \frac{e_j}{(\frac{1}{T-m})\sum_T^{t=m+1}|y_t - y_{t-m}|}$$

   The mean absolute scaled error is given by[36]

$$\text{MASE} = mean(|q_j|)$$

In addition to these performance metrics, we can also computing some statistics like correlation function and autocorrelation function.

A correlation function measures the strength of the relationship between two variables. Figure 1.12 shows different scatterplots depiction correlation between the carbon footprint and city mpg variables.

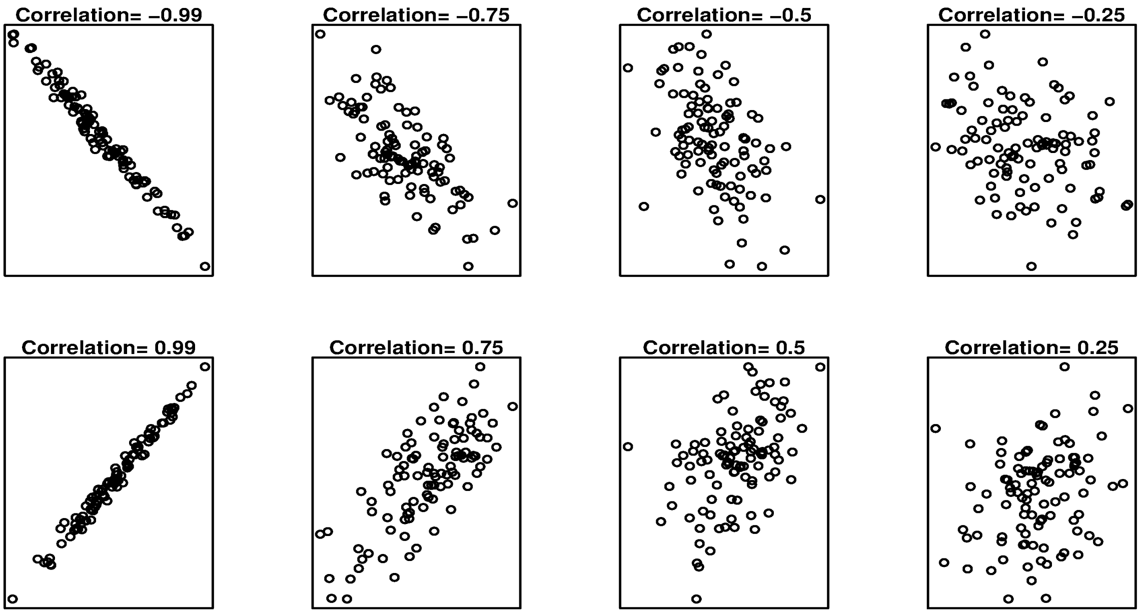

Figure 1.12: Examples of data sets with different levels of correlation

Correlation function can be written as

$$r = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2}\sqrt{\sum(y_i-\bar{y})^2}}$$

where the first variable is denoted by x and the second variable by y. The correlation coefficient only measures the strength of the linear relationship; it is possible for two variables to have a strong non-linear relationship but low correlation coefficient. The value of r always lies between -1 and 1 with negative values indicating a negative relationship and positive values indicating a positive relationship[35].

Just as correlation measures the extent of a linear relationship between two variables, autocorrelation measures the linear relationship between lagged values of a time series. There are several autocorrelation coefficients, depending on the lag length. For example, r_1 measures the relationship between y_t and y_{t-1}, r_2 measures the relationship between y_t and y_{t-2}, and so on. Figure 1.13 shows scatterplots of the beer production time series where the horizontal axis shows lagged values of the time series. Each graph shows y_t plotted against y_{t-k} for different values of k. The autocorrelations are the correlations associated with these scatterplots.

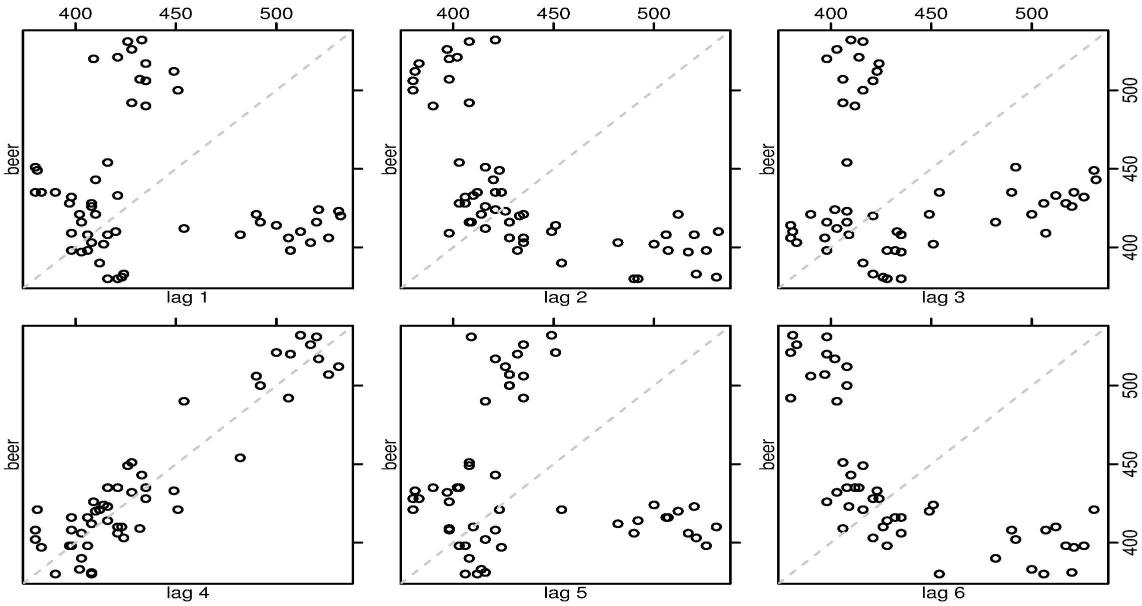

Figure 1.13: Examples of data sets with different lags of autocorrelation

The value of r_k can be written as

$$r_k = \frac{\sum_{T}^{t=k+1}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{T}^{t=1}(y_t - \bar{y})^2}$$

where T is the length of the time series[35].

CHAPTER 2

Experiments and Analysis

This chapter provides an insight into the data set being used, the tools involved in the process and the various experiments performed on the data set using the forecasting techniques described in the previous chapter.

We have divided the experiments based on the level of granularity being analysed, so we will be observing how different forecasting methods behave for different levels of granularity and meanwhile also see how any other factors in the data like trends, seasonality and non linearity influence the outcomes. For this purpose we identified certain product codes which showed some good sales patterns and hence would be helpful in analysing the models.

2.1 Data set

This research was performed on aggregated point of sales data of one particular retailer for products manufactured by a leading fast moving consumer goods company. In order to begin with the research our first step was to collect historical sales data from the particular retailer for products sold by the FMCG company, this was done using the vendor managed inventory system.

The FMCG company had access to this particular retailers POS data, and we were provided with 5 different sheets having daily POS data from 5 different store locations of the retailer, some of the data was in excel and some of them in notepad. Our next task was to identify the SKU's which were common across all these 5 stores, since it was found that some of the SKU's we not sold at all the 5 stores. Once we were able to narrow down the list of SKU's we had around 1500 SKU's, we then integrated all these SKU's in a single excel sheet and aggregated the sales of these SKU's across the 5 stores and came up with an aggregated daily sales figure for each one of them. Having the integrated data, the next task was to ensure that the format of all the attributes were the same, and here we found that there were some differences in the format of the attribute *date of sale*, especially in the way the months and the years were written, so we ensured that the date format was common. Hence we had a new data set with 1500 SKU's.

The next hurdle we came across was that the time period during which these SKU's were sold was different and some of them had really very short life and also considering the fact it would be difficult to analyse 1500 SKU's in the time period of the thesis, we considered selecting only those SKU's which were sold for the longest period. This decision was also made considering the fact that it is possible have accurate sales forecasting only for SKU's with enough past sales data, so for the SKU's which were sold for short periods it would be difficult to do so.

So we again narrowed down the list from 1500 SKU's to around 697 SKU's which had historical sales data for a period of 4 years starting from *4/2010* to *3/2014*. So we finally had a dataset of 697 different SKU's having daily aggregated POS data over a period of 4 years.

As part of data cleaning we removed some overhead information from the data set which would not be useful to us in anyway as part of this thesis, such as barcode data of the SKU, manufacturing company's code and name, retailers location, weight of the product, sub category of the SKU. Also as part of the confidentiality agreement with the industry partner we also removed the product name, as we could identify the SKU's with their unique product codes. Hence the final data set contained Product code, Date of sale, Units sold, Total sales, as seen in the dataset snippet in figure 2.1.

Product.Code	Date.of.Sale	Units.Sold	Total.Sales
228863	4/2/2010	189	1494.79
228863	4/3/2010	300	2367.25
228863	4/4/2010	303	2399.76
228863	4/5/2010	226	1789.92
228863	4/6/2010	255	2016.68
228863	4/7/2010	236	1863.28
228863	4/8/2010	222	1758.24

Figure 2.1: Snippet of the dataset

After cleaning up our data, we performed an exploratory analysis to have a better understanding of the nature of the data. This process was helpful in understanding if the SKU's consisted of any trends or seasonality and if their sales pattern was linear or non linear. We used scatterplots and time series decompostion in R in order to understand these aspects.

It was observed that though most of the SKU's showed either cyclic behaviour or had irregular patterns, there were still quite a few of them which showed some seasonality and fewer which showed trends. It was also seen that most of the SKU's had their own characteristics and specific sales pattern. Furthermore it was also observed that relationship between the SKU's was mostly non linear. Due to the strong non linear behaviour of the SKU's we were aware that techniques which model non linear behaviour would be more successful that those which model purely linear behaviour, but since our focus was mainly on how they behave while forecasting on higher granularity data i.e daily or weekly sales, it was to be seen if non linear techniques would excel under these conditions as well.

2.2 Tools used

We performed all the analysis, modeling and experimentation in *R*. It is an open source programming language which is very easy to learn and use, it is also very interactive as it helps us explore and experiment which leads to better data analysis. R scripts also help in accelerating the data analysis process, since they automates the sequence of tasks.

R also has excellent data visualization abilities which makes the process of data analysis really easy, we can create a lot of graphs like the common bar graphs and scatterplots to multi panel lattice charts or for that matter even something new that we would like to showcase. We mainly used R because of these reasons and also as it has a very strong inbuilt statistical analysis toolkit which can help in lot of tasks from data manipulation to building models to visual analysis.

It also has a lot of packages available which are free and these keep increasing by the day, and help us do almost any kind of task with real ease. Courtesy a very lively online community, R is fun to use and learn as there are a lot of materials and resources available which help us get better in using R. With the number of users and their contributions growing, R is just getting better by the hour.

2.3 Experiments based on monthly sales

To analyse the behaviour of models based on monthly sales data, we first selected product code *4661342* which is a brand of shampoo. From the initial analysis we could see some seasonality in the data, and hence we decided to check how seasonality affects the forecasting capabilities of the different models. There are 48 data points in total for monthly sales, out of which we tool 6 data points in the test set.

Figure 2.2 shows how different models have fit the monthly sales data for product code 4661342, we can see that the sales behaviour is repeating at

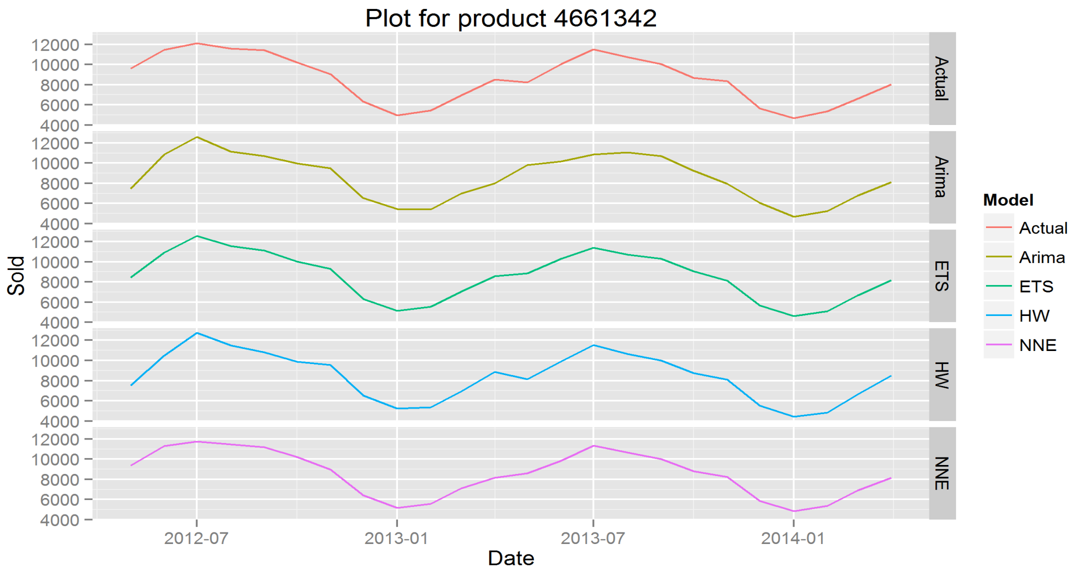

Figure 2.2: Data fit for product code 4661342 for monthly sales

regular intervals of time. We will now analyse each model step by step and explain the methodology used and how the modeling was performed.

- ARIMA

  The general approach to fit an ARIMA model to time series data is as follows :

  1. Plot the data and identify any specific observations in the data.
  2. If varaince is found to be changing over time, then stabilize it using Box-Cox transformation method.
  3. If data is found to be non stationary, then make it stationary by differencing the data.
  4. Using autocorrelation (ACF) and partial autocorrelation (PACF) find suitable AR(p) and MA(q) models.
  5. Run all the models and select the one with lowest AICc value.
  6. Calculate forecasts using the chosen model

  From figure 2.2 it can be seen that there is nothing unusual about the data, there seems to be a some seasonality due to which there is periodic rise and fall, hence it does not require any data adjustments. Also there does not seem to be any large change in variance, thus it does not require any transformation either. We also tested the same using the variance detection method using the changepoint package in R. We used two methods namely AMOC and BinSeg and they did not show any change in variance.

Though we might consider the series to might be non stationary due to the presence of seasonality, hence we performed stationarity tests to check the same. The results for the tests are as follows:

1. Augmented Dickey-Fuller Test
   H_0 : The series is non stationary
   Results :
   Dickey-Fuller = -5.845, Lag order = 3, p-value = 0.01
   alternative hypothesis: stationary
   Small p value (< 0.05), hence we can reject the null hypothesis which indicates that the series is stationary.
2. KPSS Test for Level Stationarity
   H_0 : The series is stationary
   Results:
   KPSS Level = 0.1785, Truncation lag parameter = 1, p-value = 0.1
   Large p value (> 0.05), hence we cannot reject the null hypothesis which indicates that the series is stationary.

In addition to these tests we can also see that in figure 2.3 the ACF spikes decay very fast, which is an indication that the series is stationary. For non stationary series the decay is very slow.

Since we are using auto.arima function, the choice of model will be done automatically. We decided to use auto.arima function instead of the arima function in order to keep the script automated so that the future users (the industry partner FMCG company) can use it without much trouble. But nevertheless we have explained below as to how the model selection takes place and how can we test if the selected model is optimum.

From the results it was observed that the best fit ARIMA model was *ARIMA(1,0,0)(1,0,0)[12] with non-zero mean* , which means that the model includes a non-seasonal AR(1) term, a seasonal AR(1) term, no differencing, no MA terms and the seasonal period is S = 12. This seasonal arima model can be written as :

The non-seasonal AR(1) component is $\phi(B) = 1 - \phi_1 B$

The seasonal AR(1) component is $\Phi(B^{12}) = 1 - \Phi_1 B^{12}$

The model is thus written as $(1 - \Phi_1 B^{12})(1 - \phi_1 B)(x_t - \mu) = e_t$

Let $y_t = (x_t - \mu)$

Multiplying both the AR components gives :

$$y_t = \phi_1 y_{t-1} + \Phi_1 y_{t-12} + (-\Phi_1 \phi_1) y_{t-13} + e_t$$

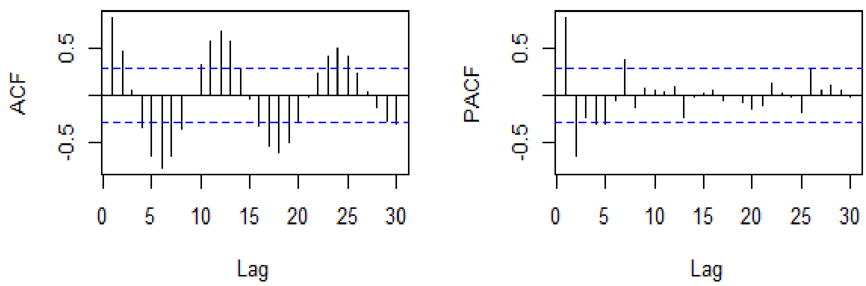

Figure 2.3: ACF and PACF plots for ARIMA model

In figure 2.3 it can clearly be seen that there is seasonality since we can see spikes at seasonal lags i.e 12, 24 etc in ACF (autocorrelation). If we now look at the PACF which is partial ACF we can see that there is a spike at lag 1 and after that there are no more significant spikes and in ACF after spike at lag 1 there is an immediate exponential decay and the same can be seen at each spike. This indicates AR(1) and since we have seasonality as well, so there would be AR(1) seasonal component as well.

The arima coefficients observed from the results are presented in table 2.1

Table 2.1: ARIMA coefficients for product code 4661342 for monthly sales

	AR1	**SAR1**	**Intercept**
Coeff	*0.7704*	*0.8849*	*0.8641*
SE Coeff	*0.0860*	*0.0464*	*0.1498*

The parameter μ is called the *intercept* in the R output.

In addition to this R also provides the value of the log likelihood of the data; that is, the logarithm of the probability of the observed data coming from the estimated model. For given values of p, d and q, R will try to maximize the log likelihood when finding parameter estimates. For our chosen model *log likelihood=-385.79*.

We also obtain values of certain indicators as specified below:

Akaike's Information Criterion (AIC), which is also useful for determining the order of an ARIMA model. It can be written as

$$AIC = -2log(L) + 2(p + q + k + 1),$$

where L is the likelihood of the data, k=1 if $c \neq 0$ and k=0 if c=0

For ARIMA models, the corrected AIC can be written as

$$AICc = AIC + 2\frac{(p+q+k+1)(p+q+k+2)}{T-p-q-k-2}$$

and the Bayesian Information Criterion can be written as

$$BIC = AIC + (log(T) - 2)(p + q + k + 1).$$

Good models are obtained by minimizing either the AIC, AICc or BIC. It has been suggested to prefer using AICc[35].

For our model the indicator values are as follows:

AIC=779.57 AICc=780.5 BIC=787.06

We can also test if this selected model has any further scope of improvement by observing the autocorrelation and performing Ljung-Box test on the forecast residuals.

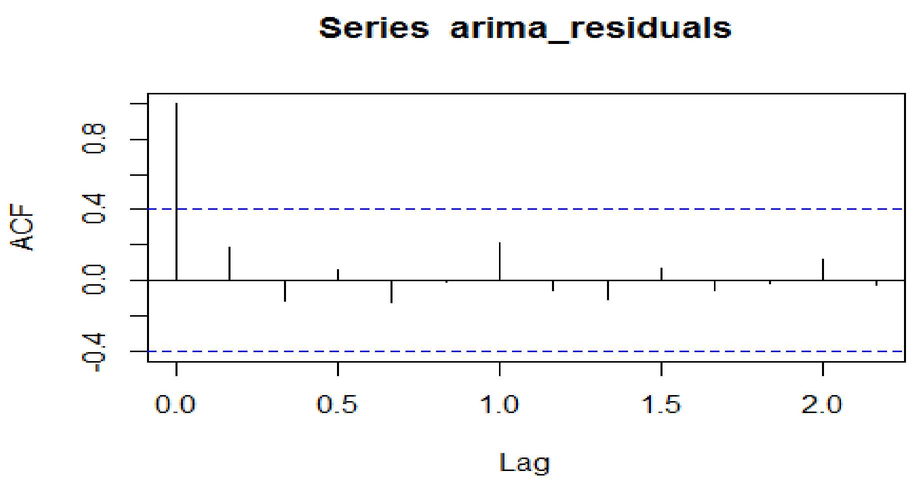

Figure 2.4: ACF of arima residuals for product code 4661342

As seen in figure 2.4 the autocorrelations for forecast errors do not exceed the threshold limits. The results of Ljung test are as follows:

$$\text{X-squared} = 13.1673, \text{ df} = 20, \text{ p-value} = 0.8701$$

A large p value indicates that the residuals are white noise and there is no further scope of model improvement.

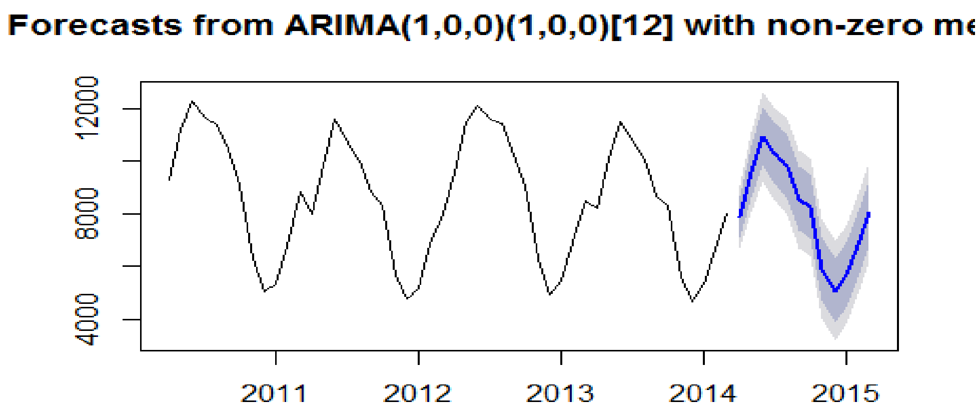

Figure 2.5: Point forecast for selected ARIMA model for product code 4661342

- Exponential smoothing (ETS)

  In this section we are using a state space model for exponential smoothing. Exponential smoothing algorithms can generate only point forecasts whereas the state space models in addition to generating point forecasts can also generate forecast intervals, hence they are more beneficial than using only exponential smoothing algorithms.

  The following steps are generally followed in order to automatically fit an ETS model to a time series:

  1. Apply all the appropriate models to the series, optimizing the parameters (both smoothing parameters and the initial state variable) of the model in each case.
  2. Select the best model among them, using information criteria, that is AIC, BIC or AICc.
  3. Produce point forecasts using the selected model with optimized parameters for the required number of steps.

4. Obtain prediction intervals for the best model. Prediction intervals can either be produced by simulation or by using algebraic formulae. If algebraic formulae is not available for some model, then for those models only simulation can be used. Simulated prediction intervals either use re-sampled errors (bootstrap method) or normally distributed errors.

ETS models are basically labelled as *Error, Trend, Seasonal.* These components can have the following possibilities :

Error (A,M), Trend =(N,A,Ad,M,Md) and Seasonal =(N,A,M)

As per the results, it was observed that for our time series data the model chosen was *ETS(M,N,M)*, which means it has multiplicative errors, no trends and multiplicative seasonality. Multiplicative models are used when data is positive, else additive models will be used. *ETS(M,N,M)* model can be written as:

$$y_t = l_{t-1} s_{t-m}(1 + \varepsilon_t)$$
$$l_t = l_{t-1}(1 + \alpha\varepsilon_t)$$
$$s_t = s_{t-m}(1 + \gamma\varepsilon_t)[37]$$

where the parameters as seen in the results are $\alpha = 0.5132$, $\gamma = 1e - 04$.

- y_t is the measurement equation which shows the relation between the observations and the unobserved state.
- l_t is the transition equation which shows the evolution of the state through time.
- ε_t is the random error process used by all the equations.
- s_t is the seasonal component

Initial values are presented in the table 2.2.

Table 2.2: Initial values for ETS model for product code 4661342

l_0	s_0	s_{-1}	s_{-2}	s_{-3}	s_{-4}	s_{-5}	s_{-6}	s_{-7}	s_{-8}	s_{-9}	s_{-10}	s_{-11}
90.44	*0.98*	*0.80*	*0.62*	*0.57*	*0.70*	*1.02*	*1.11*	*1.25*	*1.29*	*1.38*	*1.23*	*1.02*

For ETS models, Akaike's Information Criterion (AIC) is defined as,

$$AIC = -2log(L) + 2k,$$

where L is the likelihood of the model and k is the total number of parameters and initial states that have been estimated.

The AIC corrected for small sample bias (AICc) is defined as

$$AICc = AIC + \frac{2(k+1)(k+2)}{T-k}$$

and the Bayesian Information Criterion (BIC) is

$$BIC = AIC + k[log(T) - 2].[35]$$

For our model the indicator values are as follows:

AIC=762.4163 AICc=775.1436 BIC=788.6131

The lower the values the better would be the models.

Figure 2.6 shows the point forecast for the selected ETS model with 80% and 95% prediction intervals.

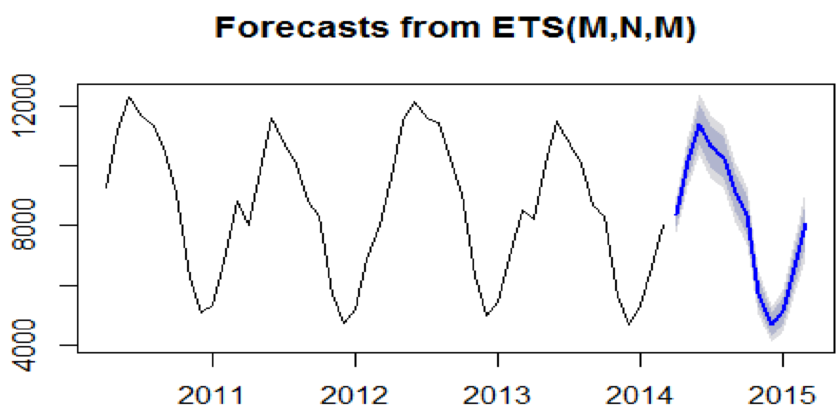

Figure 2.6: Point forecast for selected ETS model for product code 4661342

- Holt Winters (HW)

  Holt Winters method is another implementation of the exponential smoothing method. The basic difference between the two methods is the criteria that they are trying to optimize. As we have seen the *ets* function used for exponential smoothing estimates the initial states and smoothing parameters by optimizing the likelihood function, on the other hand the *holtwinters* function for the holt winters method uses a heuristic algorithm to first find the values of initial states and then estimates the smoothing parameters by optimizing the mean square prediction errors. Due to the way in which the initial states are computed it has been advised to have minimum 3 years of data to implement seasonal forecasts using the holtwinters method.

This method provides three exponential smoothing equations, one each for level which is smoothed to provide an average value to the series, trend and the seasonal component where every seasonal sub series is smoothed to provide seasonal estimate for each seasonal series. We will be using an additive holtwinters model.

This model can be written as :

$$y^{\wedge}_{t+h|t} = l_t + hb_t + s_{t-m+h_m^+}$$
$$l_t = \alpha(y_t - s_{t-m}) + (1-\alpha)(l_{t-1} + b_{t-1})$$
$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}$$
$$s_t = \gamma(y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m}[37]$$

where

l_t is the level equation which takes a weighted average of the current observation with the existing estimate of the appropriate seasonal effect subtracted, and the forecast of the level made in the earlier step.

b_t is the trend equation which takes the weighted average of the latest estimate of trend, it is given by the difference in the estimated level at time t and the estimated level at time t-1, and the previous estimate of trend.

s_t is the seasonal equation which takes the weighted average of the difference between the current observation and current estimate of the level, and the last estimate of the seasonal effect for this season which was made at time t-m.

The values of smoothing parameters as seen from our results are :

$$\alpha : 0.542187$$
$$\beta : 0$$
$$\gamma : 0.2471097$$

The values of the smoothing parameters are in the range (0,1). The above results indicate that following:

- Value of α is relatively high which means that the estimate of the level at the current point in time is based mostly on the very recent observations.
- Value of β is 0 which indicates that there is no change in the estimate of the trend component over time.
- Value of γ is quite low which means that estimate of the seasonal component at current point in time is based on some observations in distant past in addition to the recent observations.

The table 2.3 shows the coefficients for the additive holt winters model, it can be observed that though there is a huge rise and fall in the values, it is seen that the values rise and fall at constant intervals.

Table 2.3: Holt Winters coefficients for product code 4661342

a	*7883.549202*
b	*-68.6768648*
s1	*-52.61394408*
s2	*1758.034934*
s3	*3354.366243*
s4	*2592.545756*
s5	*-52.61394408*
s6	*1758.034934*
s7	*3354.366243*
s8	*-2409.529818*
s9	*-52.61394408*
s10	*1758.034934*
s11	*3354.366243*
s12	*296.216212*

We can also test if this selected model has any further scope of improvement by observing the autocorrelation and performing Ljung-Box test on the forecast residuals.

As seen in figure 2.7 the autocorrelations for forecast errors do not exceed the significance bounds for lags 1-20. The results of Ljung test are as follows:

$$\text{X-squared} = 15.3529, \text{df} = 20, \text{p-value} = 0.7559$$

Also the p-value for Ljung-Box test is 0.7, which shows that there is little chance of non-zero autocorrelations at lags 1-20.

Hence we can say that the model chosen by the holtwinters technique is suitable and there no scope for further improvement.

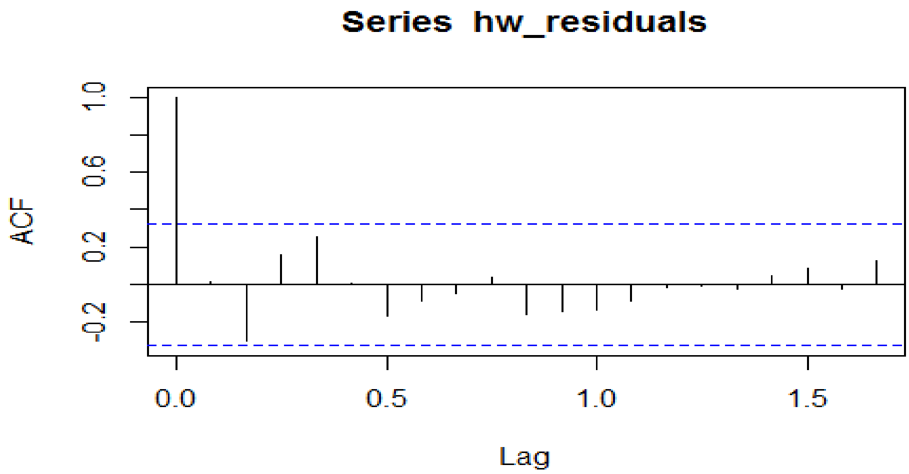

Figure 2.7: ACF of holtwinters residuals for product code 4661342

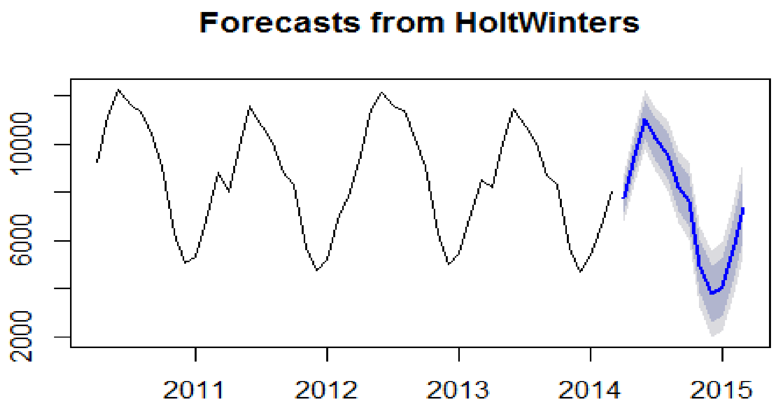

Figure 2.8: Point forecast for selected Holtwinters model for product code 4661342

Figure 2.8 shows the point forecast for the selected Holtwinters model.

- Neural Network (NNE)

  We have used nnet function to fit neural network model with lagged values of time series as inputs. It is a feed- forward network consisting of 1 hidden layer and is denoted by NNAR(p,k) where p indicated the lagged inputs and k is the number of nodes in the hidden layer.

To get the benefit of seasonal component we used the NNAR$(p, P, k)_m$ model, where the last observed values from the same season are also added to the inputs, here if the values of p and P are not specified then they are auto selected. For seasonal time series the default value of P is 1 and p is selected from optimal linear model fitted to the seasonally adjusted data, which in our case will be 13. For non seasonal time series the default value of p is the optimum number of lags as specified by AIC for a linear AR(p) model. The value of k is also auto selected if unspecified, it is set to $k = \frac{(p+P+1)}{2}$ (rounded to the nearest integer). A NNAR$(p, P, 0)_m$ model has inputs $(y_{t-1}, y_{t-2}, \ldots, y_{t-p}, y_{t-m}, y_{t-2m}, y_{t-Pm})$ and k neurons in the hidden layer. It is equivalent to an ARIMA $(p, 0, 0)(P, 0, 0)_m$ model but without the restrictions on the parameters to ensure stationarity [35].Thus we use a 13-8-1 network for our modelling of seasonal data using neural network. It is a NNAR$(13, 1, 8)_{12}$ model and has inputs $y_{t-1}, y_{t-2}, \ldots, y_{t-13}$ and 8 neurons in the hidden layer. In short the feed-forward neural network is fitted with lagged values of the given time series as inputs and a single hidden layer where the nodes are either specified or selected by default. The inputs are for lags 1 to p, and lags m to mP where m refers to the frequency of the time series. A total of 20 to 30 networks are fitted unless specified, each with random starting weights. These are then averaged to compute the forecasts.

From the results it is seen that an average of 20 networks each of which is a 13-8-1 network are trained and their predictions averaged and the output units are kept linear.

One downside of this package is that it can only compute point forecasts and not forecast intervals. This can be seen in figure 2.9.

Also unlike other libraries in the forecasting package which allowed a lot of visualization to be performed, we found that the *nnet* library for applying time series to neural networks was not very flexible in that aspect and hence we had some difficulty in analysing the results at the first instance and we really had to break up the code manually to understand what is happening at each stage and how the network is evolving. Also in literature neural networks in R have been criticized as black boxes which do not offer a lot of information regarding the relation between variables. This can be considered as some sort of a limitation for time series neural network application in R.

The figure 2.10 shows the forecasts of all the four models discussed above.

We will now see how all the models performed by using the accuracy measures obtained from the results. The results are shown in table 2.4.

Overall it is observed that the neural network model behaves the best followed by exponential smoothing, holt winters and ARIMA. This was expected since the data is highly non linear and ARIMA does not fit well for non linear

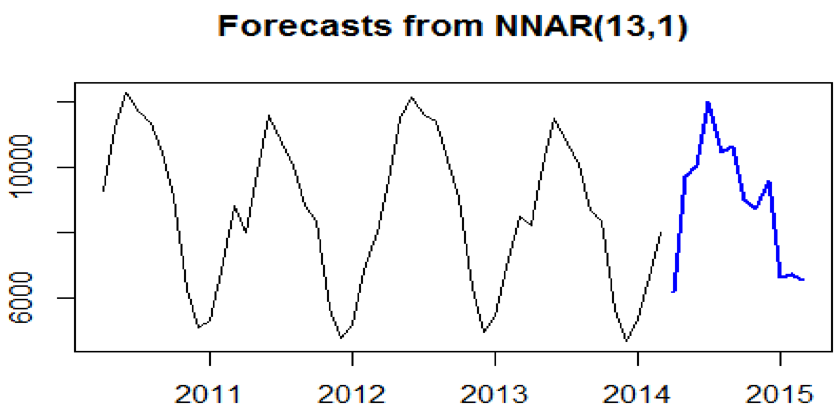

Figure 2.9: Point forecast for neural network model for product code 4661342

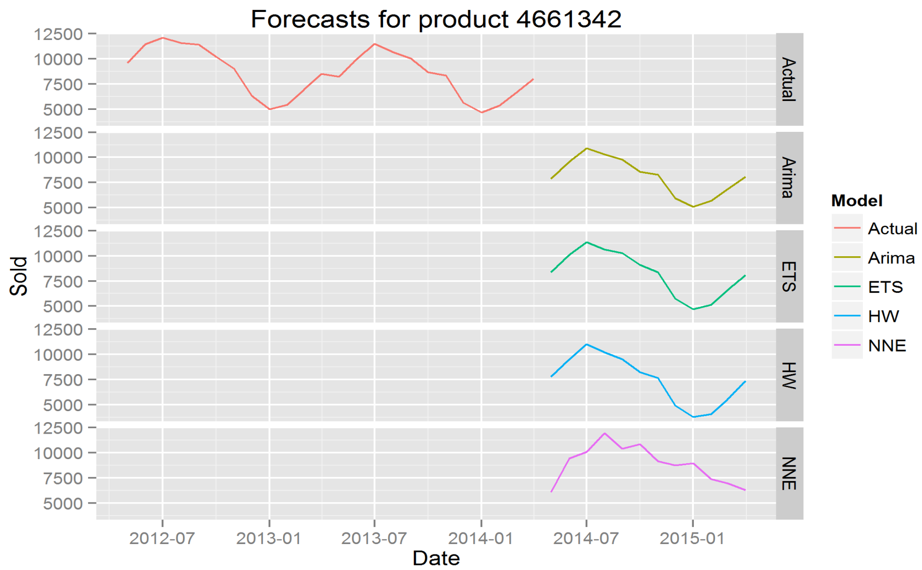

Figure 2.10: Forecasts for product code 4661342

data whereas neural networks work the best and it is also seen that seasonality also has been taken care well by the neural network model. Another striking factor is that the exponential smoothing model is working almost as good as the neural network model if we consider the MAPE values, MAPE and MASE are generally the preferred accuracy measures. Due to usage of

Table 2.4: Accuracy measures for product code 4661342

Model	RMSE	MAE	MAPE	MASE
Arima	*612.207*	*461.971*	*5.509*	*0.382*
NNE	*180.543*	*153.977*	*2.0236*	*0.127*
ETS	*321.717*	*226.889*	*2.651*	*0.187*
HW	*491.177*	*314.770*	*3.989*	*0.255*

aggregated sales figures it is better not consider RMSE values.

We next experimented another product code which had some trend in its behaviour. It is a brand of pet food, with product code *814593*. Figure 2.11 shows how different models have fit the monthly sales data for product code 814593, we can see an increasing trend in the sales behaviour.

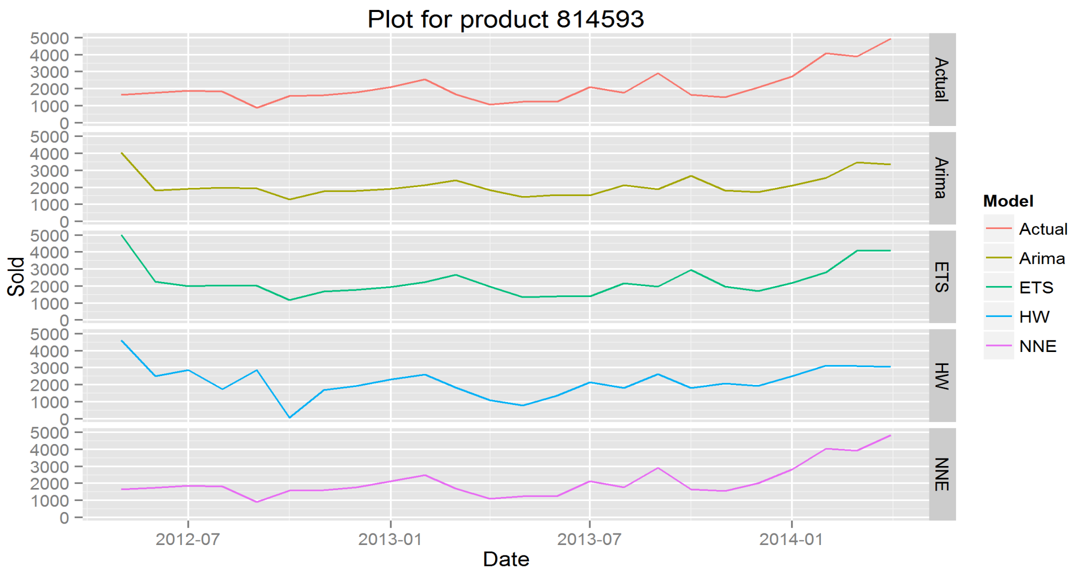

Figure 2.11: Data fit for product code 814593 for monthly sales

From our experimental results we saw that it uses *ARIMA(1,0,0) with non-zero mean model*, which means that it includes a non seasonal AR(1) term and since we cannot see any seasonality in the pattern we do not have the seasonal component. The indicator values were seen as

AIC=779.25 AICc=779.79 BIC=784.86

The arima coefficients observed from the results are presented in table 2.5.

Table 2.5: ARIMA coefficients for product code 814593 for monthly sales

	AR1	**Intercept**
Coeff	*0.6746*	*21.97*
SE Coeff	*0.1201*	*3.25*

The ETS model selected was *ETS(M,A,N)* which means it has multiplicative errors, additive trends and no seasonality. This model can be written as

$$y_t = (l_{t-1} + b_{t-1})(1 + \varepsilon_t)$$
$$l_t = (l_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$$
$$b_t = s_{t-1} + \beta((l_{t-1} + b_{t-1}))\varepsilon_t[37]$$

where the parameters as seen in the results are $\alpha = 0.867$, $\beta = 1e - 04$. Initial values are given as $l_0 = 1405.7456$ and $b_0 = 156.8864$.

The indicator values shown were :

AIC=812.37 AICc=813.30 BIC=819.85

The holtwinters makes use of an additive model and shows the following values for the smoothing parameters :

$$\alpha : 0.7269292$$
$$\beta : 0$$
$$\gamma : 0.3546306$$

It can be seen that the holtwinters model does not detect any trend in the data whereas the exponential function seen earlier had a trend component. This is might be a result of the way these methods are optimizing the smoothing parameters.

The neural network model again makes use of an average of 20 networks, each of which is a 13-8-1 network with linear output units.

We will now see how all the models performed by using the accuracy measures obtained from the results. The results are shown in table 2.6.

Table 2.6: Accuracy measures for product code 814593 for monthly sales

Model	**RMSE**	**MAE**	**MAPE**	**MASE**
Arima	*615.227*	*463.771*	*5.455*	*0.372*
NNE	*118.549*	*87.322*	*1.104*	*0.072*
ETS	*318.727*	*212.789*	*2.645*	*0.187*
HW	*491.177*	*311.670*	*3.999*	*0.276*

If we look at the MAPE values we can see that neural networks again outperforms all the other models and the margin between neural networks and exponential smoothing though not very large is comparatively bigger than the previous case. ARIMA continues to perform badly because of the non linearity in the data.

The figure 2.12 shows the forecasts of all the four models discussed above.

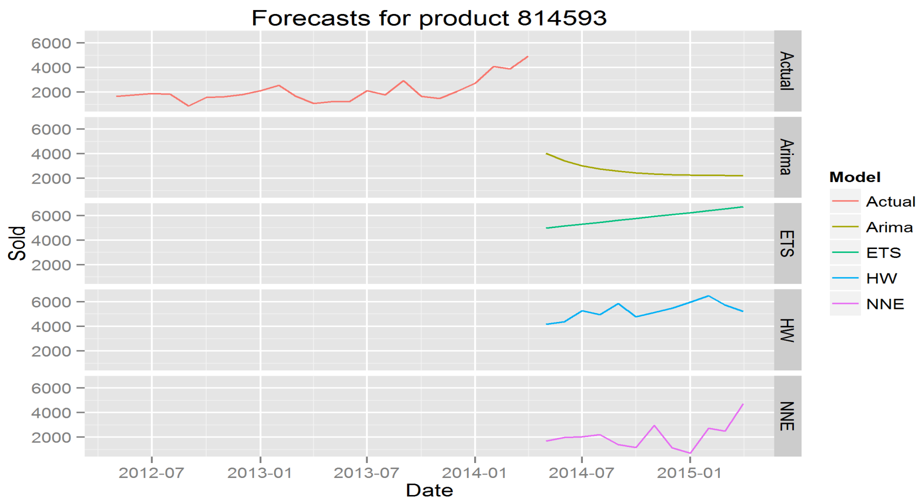

Figure 2.12: Forecasts for product code 814593 for monthly sales

In the next experimented we used another product code which showed very irregular behaviour, the sales did not show any sort of pattern. It is a brand of women care, with product code *4610600*. Figure 2.13 shows how different models have fit the monthly sales data for product code 4610600.

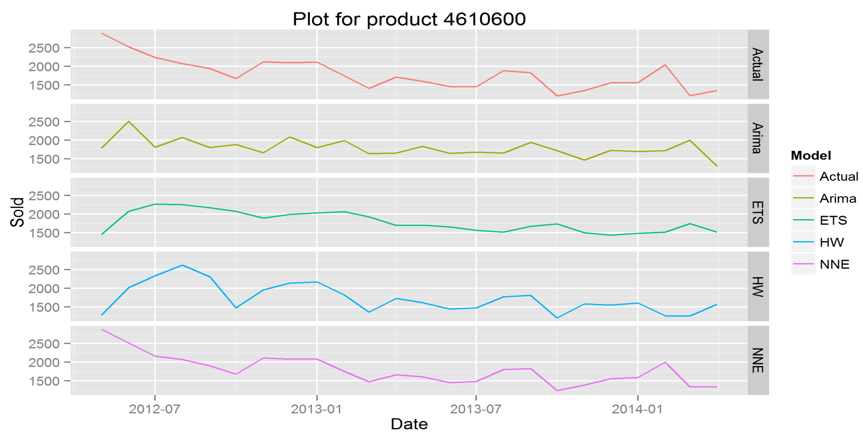

Figure 2.13: Data fit for product code 4610600 for monthly sales

From our experimental results we saw that it uses *ARIMA(0,0,1) with non-zero mean*, which means that it includes a non seasonal MA(1) term and since we cannot see any seasonality in the pattern we do not have the seasonal component. MA (1) means that the lag-1 autocorrelation is negative and the ACF cuts off beyond lag 1.

$$y_t = \theta_1 e_{t-1}$$

The arima coefficients observed from the results are presented in table 2.7.

Table 2.7: ARIMA coefficients for product code 4610600 for monthly sales

	MA1	**Intercept**
Coeff	*0.6351*	*18.01*
SE Coeff	*0.1247*	*7.94*

The indicator values were seen as

AIC=702.03 AICc=702.57 BIC=707.64

The ETS model selected was *ETS(A,N,N)* which means it has additive errors, no trends and no seasonality. This model can be written as

$$y_t = l_{t-1} + \varepsilon_t$$
$$l_t = l_{t-1} + \alpha\varepsilon_t$$

where the parameters as seen in the results are $\alpha = 0.4327$ and initial values $l_0 = 2506.3269$.

The indicator values shown were :

AIC=754.5818 AICc=754.8485 BIC=758.3242

The holtwinters makes use of an additive model and shows the following values for the smoothing parameters :

$$\alpha : 0.5186252$$
$$\beta : 0$$
$$\gamma : 0.1123041$$

The neural network model again makes use of an average of 20 networks, each of which is a 13-8-1 network with linear output units.

The figure 2.14 shows the forecasts of all the four models discussed above.

We will now see how all the models performed by using the accuracy measures obtained from the results. The results are shown in table 2.8.

If we look at the MAPE values we can see that neural networks again outperforms all the other models followed by holtwinters, ARIMA and exponential smoothing. The sales for this particular product code really had no

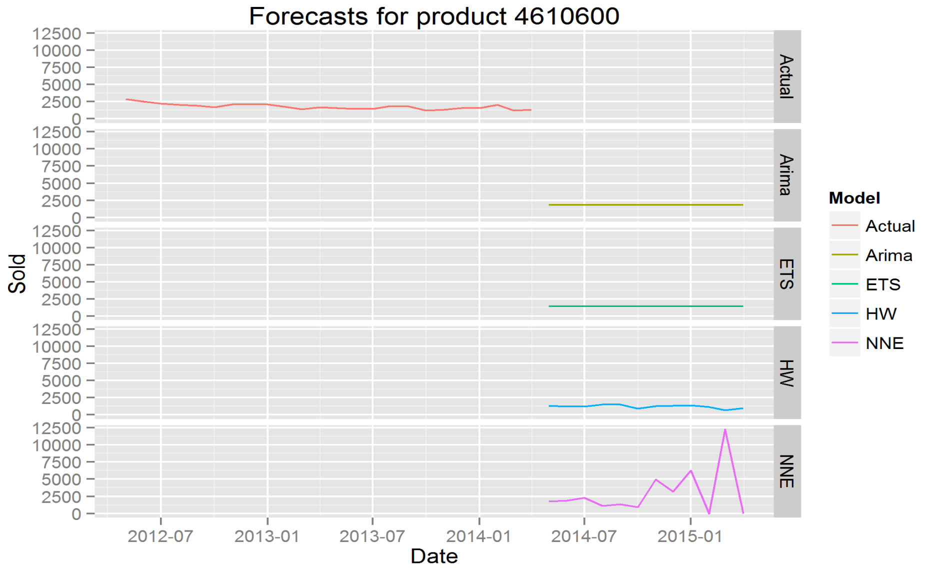

Figure 2.14: Forecasts for product code 4610600 for monthly sales

Table 2.8: Accuracy measures for product code 4610600 for monthly sales

Model	RMSE	MAE	MAPE	MASE
Arima	*338.982*	*253.82*	*14.753*	*0.916*
NNE	*32.314*	*22.667*	*1.466*	*0.081*
ETS	*358.869*	*275.206*	*15.971*	*0.994*
HW	*370.485*	*198.207*	*11.636*	*0.706*

specific pattern which makes forecasting all the more difficult and this can be seen in figure 2.14. Forecasts made by ARIMA and exponential smoothing models is difficult to interpret as they seem to have smoothed out.

We can also see that the MAPE values have increased for all the models compared to the previous experiments and though neural networks still manages to perform well, all the other models seem to be way behind. This clearly indicated that though it is rather difficult to forecast sales which have no specific pattern, neural network still manages to do so, but the traditional methods suffer.

We also observed that there was no possibility to generalise the outcomes by categorizing the SKU's as per their sub category or brand or family, since while experimenting it was found that the behaviour is really SKU specific and each SKU in a product family had a different pattern and hence the results varied accordingly. This makes it difficult to say that this method works well for baby care products or some other method works well for pet foods.

2.4 Experiments based on weekly sales

In order to maintain continuity in the writeup and to compare the outcomes, we will choose the same products that were used in our earlier experiment involving monthly sales data.

We will first check the product code *4661342*. Figure 2.15 shows how different models have fit the weekly sales data for product code 4661342.

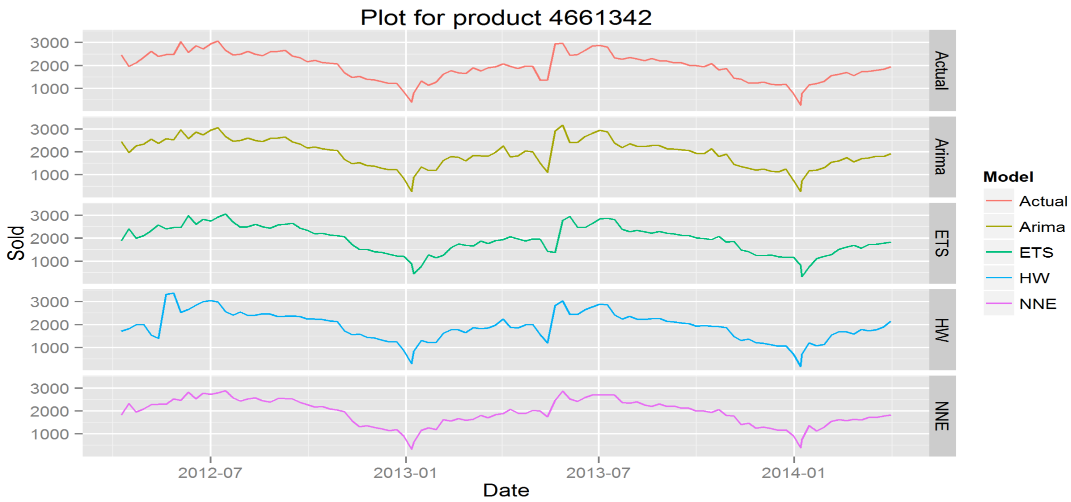

Figure 2.15: Data fit for product code 4661342 for weekly sales

From the results we can see that *ARIMA(1,1,0)(1,1,0)[53]* is used, which means that it includes a non seasonal AR(1) term and simple differencing of order 1 and due to weekly seasonality we even have a seasonal component having seasonal AR(1) and seasonal differencing of order 1.

From the figure 2.16 we can see that ACF lag at 1 is not dying out quickly which shows that the series is non stationary and hence needs differencing.

We then apply differencing of order 1 to the time series to make it stationary as seen in figure 2.17.

The indicators are as below:

AIC=1111.76 AICc=1111.92 BIC=1120.95

The arima coefficients observed from the results are presented in table 2.9.

Table 2.9: ARIMA coefficients for product code 4661342 for weekly sales

	AR1	**SAR1**
Coeff	*-0.4195*	*-0.9423*
SE Coeff	*0.0644*	*0.0130*

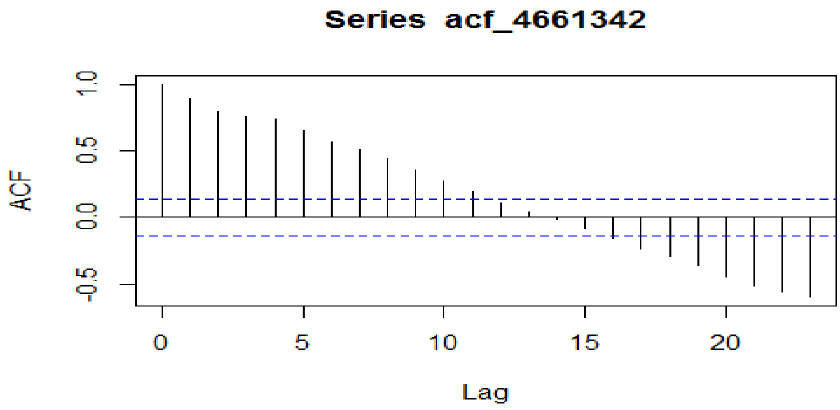

Figure 2.16: ACF of product 4661342 for weekly sales

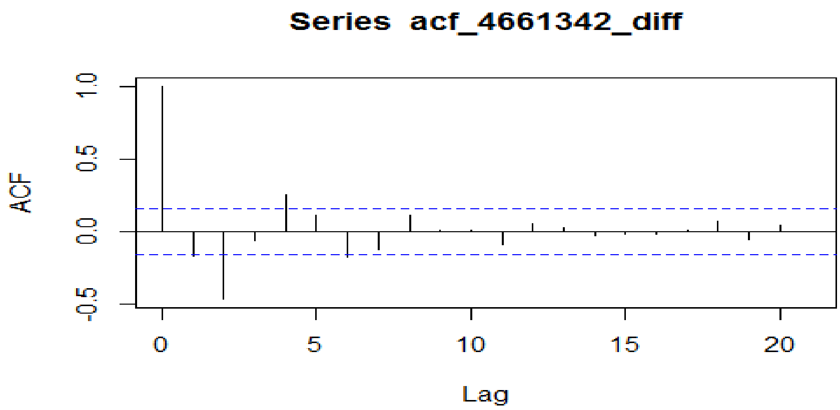

Figure 2.17: ACF of product 4661342 for weekly sales after differencing

The ETS model selected was *ETS(A,N,N)* which means it has additive errors, no trends and no seasonality. This model can be written as

$$y_t = l_{t-1} + \varepsilon_t$$
$$l_t = l_{t-1} + \alpha\varepsilon_t$$

where the parameters as seen in the results are $\alpha = 0.9019$ and initial values $l_0 = 2126.0213$. T he indicator values shown were :

AIC=3528.968 AICc=3529.026 BIC=3535.682

The holtwinters makes use of an additive model and shows the following values for the smoothing parameters :

$$\alpha : 0.1520057$$
$$\beta : 0$$
$$\gamma : 0.1060408$$

The parameters are both very small, hence it means that means that estimate of the seasonal component at current point in time is based mostly on the past observations.

The neural network model makes use of an average of 22 networks, each of which is a 5-3-1 network with linear output units and with 22 weights.

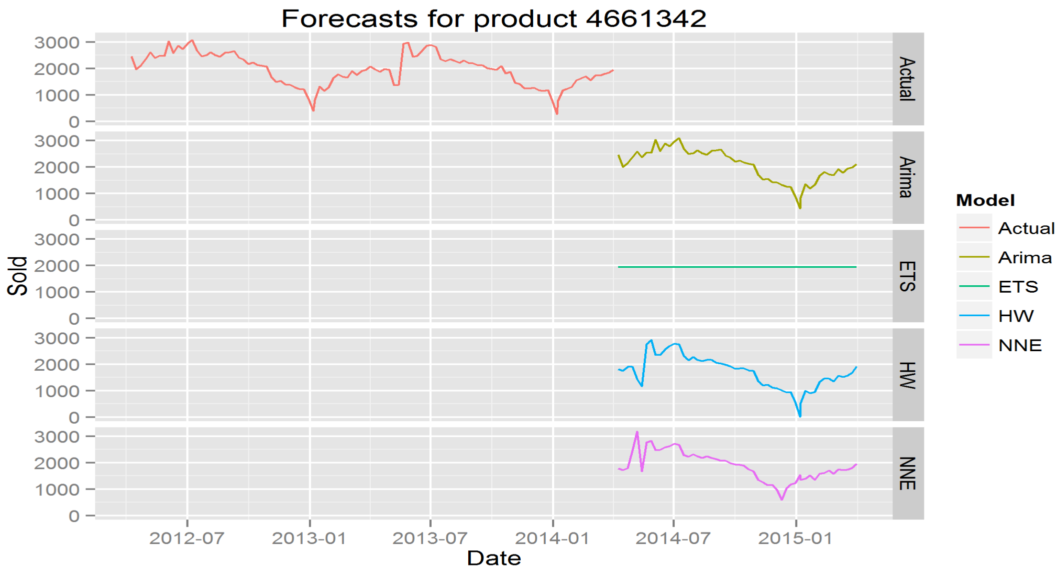

Figure 2.18: Forecasts for product code 4661342 for weekly sales

We will now see how all the models performed by using the accuracy measures obtained from the results. The results are shown in table 2.10.

Table 2.10: Accuracy measures for product code 4661342 for weekly sales

Model	**RMSE**	**MAE**	**MAPE**	**MASE**
Arima	*41.386*	*18.672*	*1.172*	*0.10165*
NNE	*180.587*	*128.722*	*7.527*	*0.70077*
ETS	*280.146*	*184.838*	*13.528*	*1.00626*
HW	*190.405*	*96.641*	*5.931*	*0.5133*

The performance metrics show that ARIMA performed the best followed by holtwinters, neural networks and exponential smoothing. ARIMA is known to perform well in case of short term forecasts, where as we can see in figure

2.18 which shows the forecasts of all the four models that the exponential smoothing method does not perform well.

We will next see how the weekly forecasts work for product code *814593*. Figure 2.19 shows how different models have fit the weekly sales data for product code 814593.

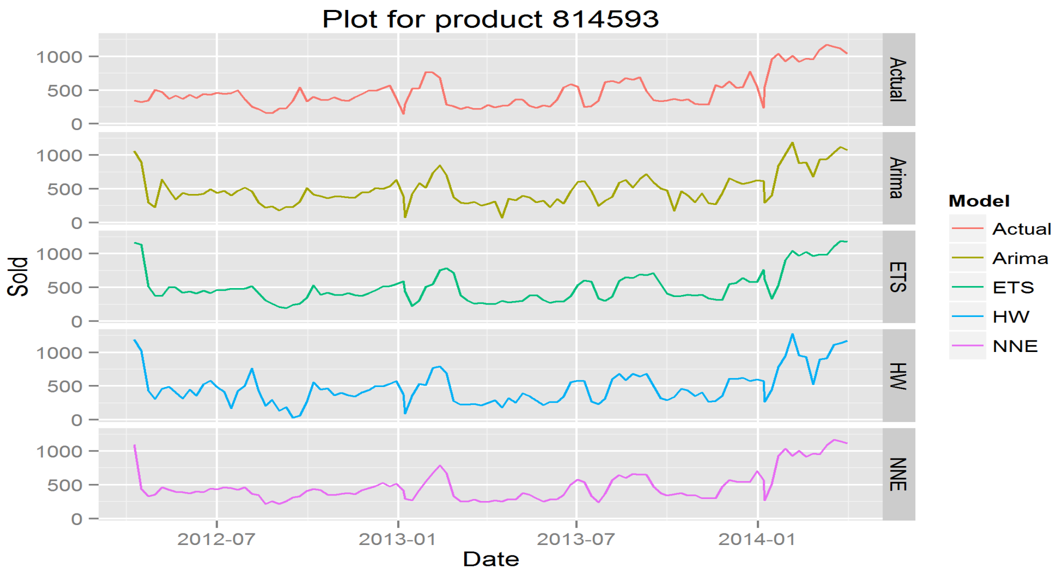

Figure 2.19: Data fit for product code 814593 for weekly sales

From our experimental results we saw that it uses *ARIMA(0,1,4)(1,0,1)[53]* , which means that it includes a non seasonal MA(4) term and a simple differencing of order 1 along with a seasonal AR(1) and MA(1) terms. The model has one differencing and there is a negative spike in the ACF at lag 4 and also a negative spike in the ACF at lag 53, whereas the PACF has a more gradual decay pattern in the vicinity of both these lags.

The indicator values were seen as

$$AIC{=}2603.17 \; AICc{=}2603.17 \; BIC{=}2626.63$$

The arima coefficients observed from the results are presented in table 2.11.

Table 2.11: ARIMA coefficients for product code 814593 for weekly sales

	MA1	**MA2**	**MA3**	**MA4**	**SAR1**	**SMA1**
Coeff	*0.1113*	*0.0166*	*-0.3440*	*-0.1272*	*0.8550*	*-0.6831*
SE Coeff	*0.0678*	*0.0726*	*0.0652*	*0.0854*	*0.0538*	*0.0676*

The ETS model selected was *ETS(M,A,N)* which means it has multiplicative errors, additive trends and no seasonality. This model can be written as

$$y_t = (l_{t-1} + b_{t-1})(1 + \varepsilon_t)$$
$$l_t = (l_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$$
$$b_t = s_{t-1} + \beta((l_{t-1} + b_{t-1}))\varepsilon_t[37]$$

where the parameters as seen in the results are $\alpha = 0.8194$, $\beta = 1e - 04$. Initial values are given as $l_0 = 309.9473$ and $b_0 = 24.2002$.

The indicator values shown were :

AIC=3205.792 AICc=3205.985 BIC=3219.218

The holtwinters makes use of an additive model and shows the following values for the smoothing parameters :

$$\alpha : 0.9672376$$
$$\beta : 0$$
$$\gamma : 1$$

Here we can see that the value of γ is 1 and α is almost equal to 1, which means that the estimate makes use of mostly recent observations.

The neural network model makes use of an average of 20 networks, each of which is a 18-10-1 network with linear output units.

We will now see how all the models performed by using the accuracy measures obtained from the results. The results are shown in table 2.12.

Table 2.12: Accuracy measures for product code 814593 for weekly sales

Model	**RMSE**	**MAE**	**MAPE**	**MASE**
Arima	*98.761*	*56.535*	*12.772*	*0.647*
NNE	*32.153*	*22.099*	*5.905*	*0.253*
ETS	*139.598*	*92.048*	*25.720*	*1.054*
HW	*119.216*	*65.768*	*19.060*	*0.766*

Compared to the previous experiment on weekly data, here we see that the neural network model has performed the best followed by ARIMA, holtwinters and again exponential smoothing seems to the worst. We have seen in literature survey that generally exponential smoothing works well with yearly data, and we also saw that it performed well in case of monthly data. Thus this shows that exponential smoothing might not be a really good method for short term forecasts.

The figure 2.20 shows the forecasts of all the four models discussed above.

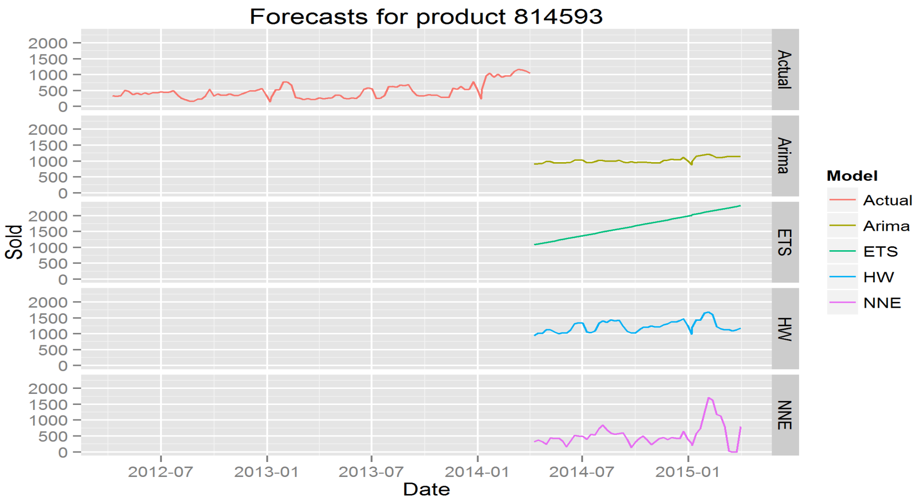

Figure 2.20: Forecasts for product code 814593 for weekly sales

Next we experimented with product code *4610600*. Figure 2.21 shows how different models have fit the weekly sales data for product code 4610600.

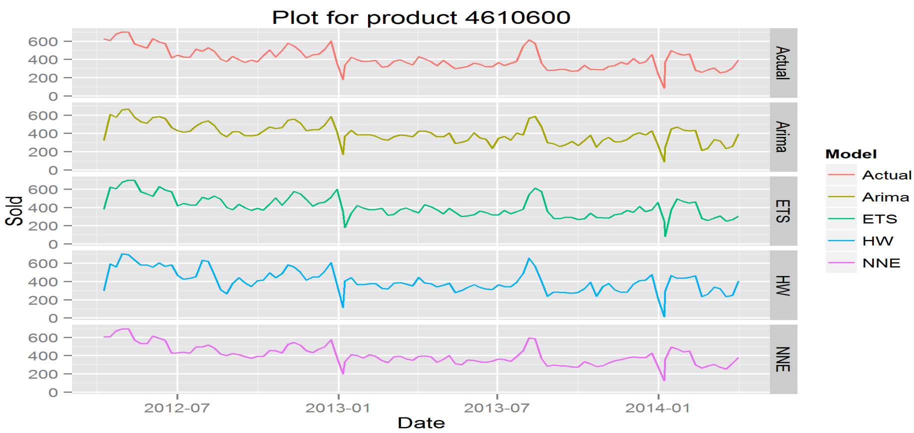

Figure 2.21: Data fit for product code 4610600 for weekly sales

From our experimental results we saw that it uses *ARIMA(3,1,1)(1,0,1)[53]*, which means that it has differencing of order with a non seasonal AR(3) and MA(1) and also a seasonal AR(1) and MA(1).

The arima coefficients observed from the results are presented in table 2.13.

Table 2.13: ARIMA coefficients for product code 4610600 for weekly sales

	AR1	**AR2**	**AR3**	**MA1**	**SAR1**	**SMA1**
Coeff	*0.8906*	*0.0679*	*-0.1920*	*-0.9227*	*0.8842*	*-0.6912*
SE Coeff	*0.0890*	*0.0898*	*0.0719*	*0.0551*	*0.0305*	*0.0629*

The indicator values were seen as

AIC=2275.17 AICc=2275.73 BIC=2298.64

The ETS model selected was *ETS(A,N,N)* which means it has additive errors, no trends and no seasonality. This model can be written as

$$y_t = l_{t-1} + \varepsilon_t$$
$$l_t = l_{t-1} + \alpha\varepsilon_t$$

where the parameters as seen in the results are $\alpha = 0.9999$ and initial values $l_0 = 575.2468$.

The indicator values shown were :

AIC=2996.615 AICc=2996.672 BIC=3003.328

The holtwinters makes use of an additive model and shows the following values for the smoothing parameters :

$$\alpha : 0.8897354$$
$$\beta : 0$$
$$\gamma : 1$$

The neural network model again makes use of an average of 20 networks, each of which is a 19-10-1 network with linear output units.

We will now see how all the models performed by using the accuracy measures obtained from the results. The results are shown in table 2.14.

Table 2.14: Accuracy measures for product code 4610600 for weekly sales

Model	**RMSE**	**MAE**	**MAPE**	**MASE**
Arima	*44.856*	*25.945*	*7.113*	*0.463*
NNE	*22.753*	*17.120*	*4.886*	*0.305*
ETS	*79.818*	*55.718*	*16.712*	*0.995*
HW	*52.812*	*32.282*	*8.853*	*0.568*

Neural network continues to perform the best followed by ARIMA, as seen in the previous experiments on weekly data we can say that ARIMA is a close contender to neural networks model while forecasting weekly data. But again we see that exponential smoothing method continue to perform badly and the same can be seen in figure 2.22. Hence it is really not recommended for weekly forecasting.

The figure 2.22 shows the forecasts of all the four models discussed above.

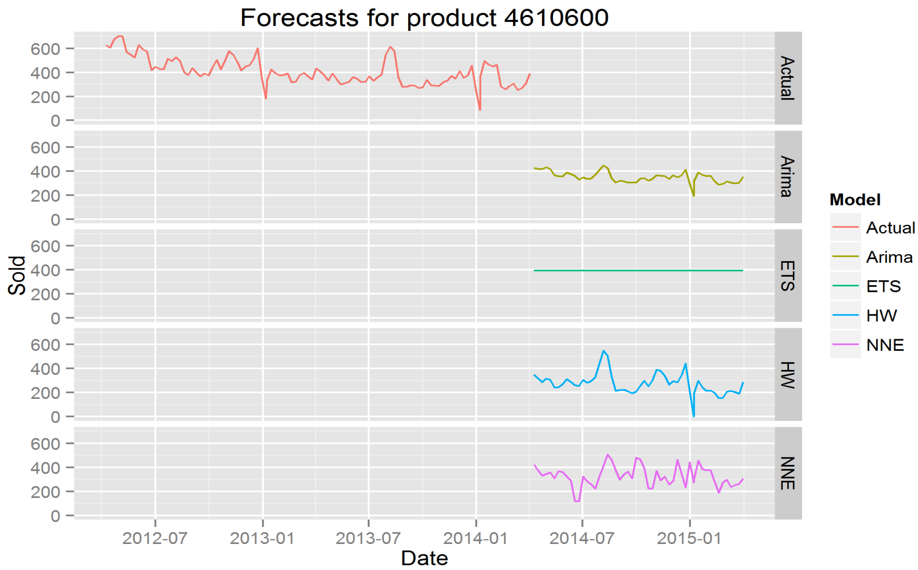

Figure 2.22: Forecasts for product code 4610600 for weekly sales

2.5 Experiments based on daily sales

Next lets see how the models behave in case of daily sales. We will begin again with product code *4661342*. Figure 2.23 shows how different models have fit the daily sales data for product code 4661342.

From the results we can see that *ARIMA(5,1,2)* is used, which means that it includes a non seasonal AR(1) term with simple differencing of order 1 and a non seasonal MA(1) term. There is no seasonal component as the daily sales data does no show any seasonality.

The indicators are as below:

AIC=15446.55 AICc=15446.65 BIC=15488.83

The arima coefficients observed from the results are presented in table 2.15.

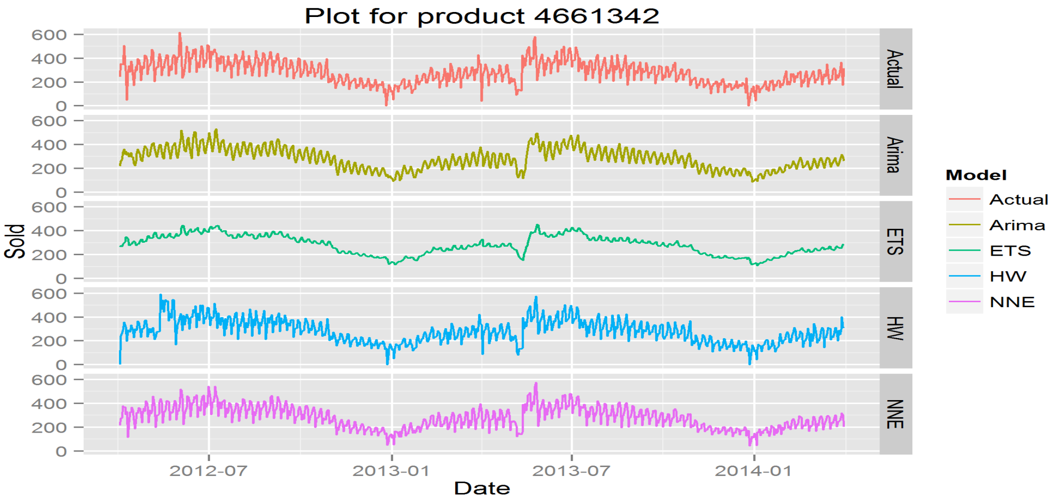

Figure 2.23: Data fit for product code 4661342 for daily sales

Table 2.15: ARIMA coefficients for product code 4661342 for daily sales

	AR1	**AR2**	**AR3**	**AR4**	**AR5**	**MA1**	**MA2**
Coeff	*0.4481*	*-0.5632*	*-0.3430*	*-0.2443*	*-0.2527*	*-1.1585*	*0.8503*
SE Coeff	*0.0322*	*0.0281*	*0.0298*	*0.0280*	*0.0285*	*0.0205*	*0.0223*

The ETS model selected was *ETS(A,N,N)* which means it has additive errors, no trends and no seasonality. This model can be written as

$$y_t = l_{t-1} + \varepsilon_t$$
$$l_t = l_{t-1} + \alpha\varepsilon_t$$

where the parameters as seen in the results are $\alpha = 0.1977$ and initial values $l_0 = 318.2791$. T he indicator values shown were :

AIC=22696.09 AICc=22696.10 BIC=22706.67

The holtwinters makes use of an additive model and shows the following values for the smoothing parameters :

$$\alpha : 0.1157$$
$$\beta : 0$$
$$\gamma : 0.1513$$

The parameters are both very small, hence it means that means that estimate of the seasonal component at current point in time is based mostly on the past observations.

The neural network model makes use of an average of 20 networks, each of which is a 30-6-1 network with linear output units and with 512 weights.

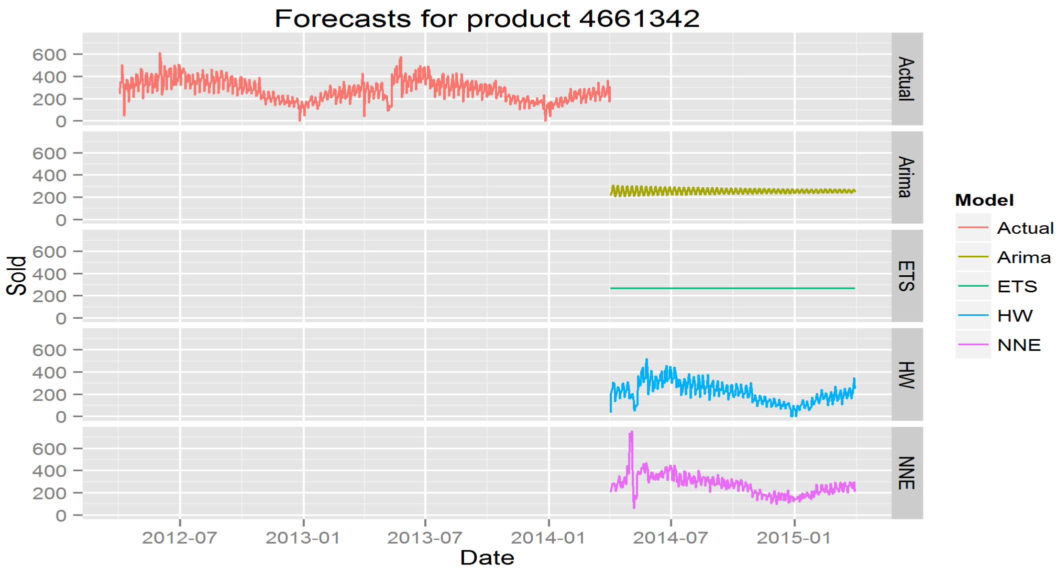

Figure 2.24: Forecasts for product code 4661342 for daily sales

We will now see how all the models performed by using the accuracy measures obtained from the results. The results are shown in table 2.16.

Table 2.16: Accuracy measures for product code 4661342 for daily sales

Model	RMSE	MAE	MAPE	MASE
Arima	*47.985*	*34.540*	*Inf*	*0.695*
NNE	*22.979*	*17.073*	*Inf*	*0.343*
ETS	*62.417*	*46.063*	*Inf*	*0.927*
HW	*48.165*	*27.738*	*Inf*	*0.552*

We see that the MAPE value are Inf, now this happens if there are zero values, which does happen in case of sales figures and especially when we are dealing with data at very high level of granularity. In this case there will be a division by zero which leads to Inf values. Hence we will consider the MASE values to evaluate the models, we see that again neural networks performs the best followed by holtwinters, ARIMA and exponential smoothing. Figure 2.24 shows that exponential smoothing still continues to perform poorly.

We will next see how the daily forecasts work for product code *814593*. Figure 2.25 shows how different models have fit the daily sales data for product code 814593.

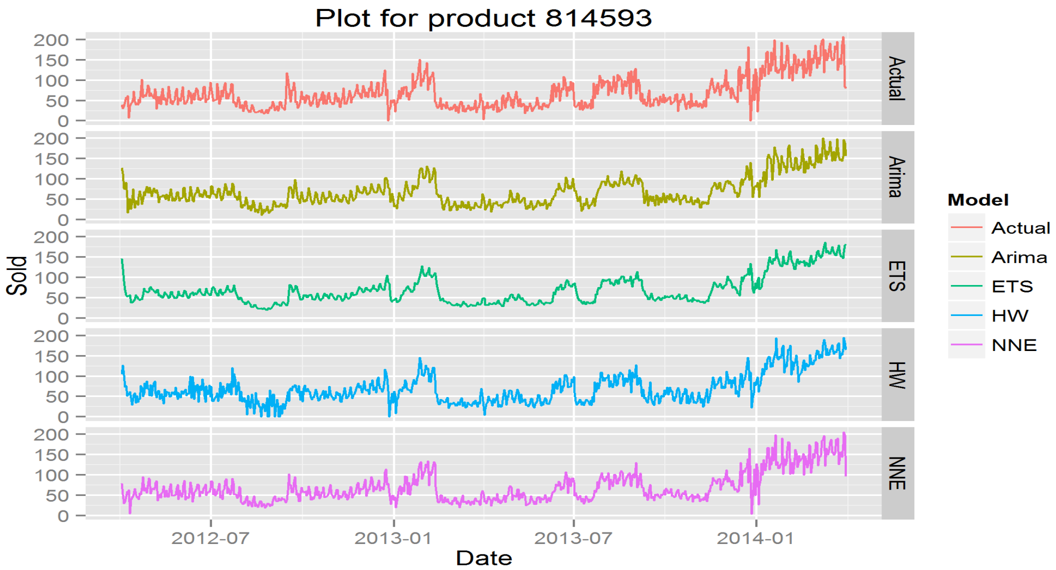

Figure 2.25: Data fit for product code 814593 for daily sales

From our experimental results we saw that it uses *ARIMA(5,1,5)* , which means that it includes a non seasonal AR(5) term and a simple differencing of order 1 along with a non seasonal AR(5). The model this has one differencing and there is a negative spike in the ACF at lag 5 and the PACF has a gradual decay pattern near this lag also there is a positive spike in PACF at lag 5 and the ACF shows a decay pattern at this lag.

The indicator values were seen as

AIC=12443.47 AICc=12443.65 BIC=12501.6

The arima coefficients observed from the results are presented in table 2.17.

Table 2.17: ARIMA coefficients for product code 814593 for daily sales

	AR1	**AR2**	**AR3**	**AR4**	**AR5**
Coeff	*0.544*	*-1.231*	*0.437*	*-0.783*	*-0.238*
SE Coeff	*0.062*	*0.051*	*0.088*	*0.051*	*0.059*
	MA1	**MA2**	**MA3**	**MA4**	**MA5**
Coeff	*-1.107*	*1.569*	*-1.193*	*1.080*	*-0.322*
SE Coeff	*0.063*	*0.063*	*- 0.089*	*0.064*	*0.055*

The ETS model selected was *ETS(A,N,N)* which means it has additive errors, no trends and no seasonality. This model can be written as

$$y_t = l_{t-1} + \varepsilon_t$$
$$l_t = l_{t-1} + \alpha\varepsilon_t$$

where the parameters as seen in the results are $\alpha = 0.3718$ and initial values $l_0 = 39.7997$.

The indicator values shown were :

AIC=19415.25 AICc=19415.26 BIC=19425.83

The holtwinters makes use of an additive model and shows the following values for the smoothing parameters :

$$\alpha : 0.3282$$
$$\beta : 0$$
$$\gamma : 0.3379$$

Here we can see that the value of α and γ are quite small, which means that the estimate makes use of mostly past observations, and some recent observation.

The neural network model makes use of an average of 20 networks, each of which is a 29-15-1 network with linear output units.

The figure 2.26 shows the forecasts of all the four models discussed above.

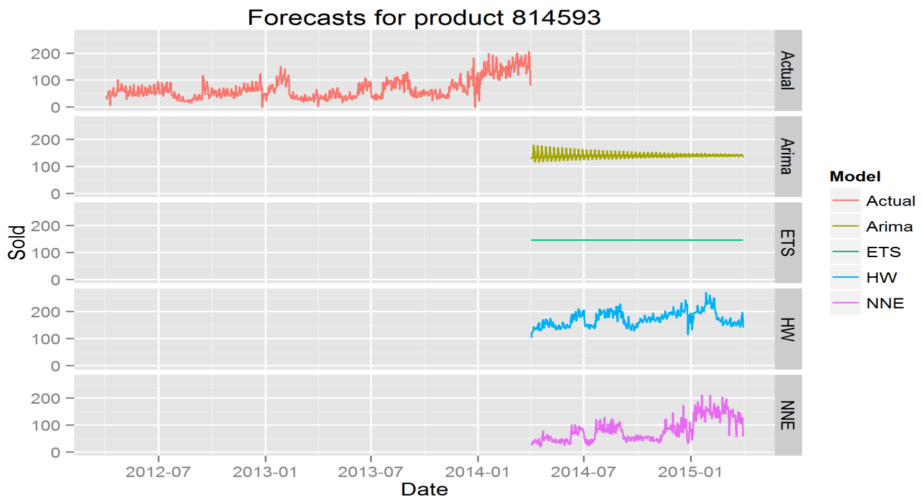

Figure 2.26: Forecasts for product code 814593 for daily sales

We will now see how all the models performed by using the accuracy measures obtained from the results. The results are shown in table 2.18.

Neural network continues to perform the best, followed by ARIMA and holtwinters. But we still see that the performance of exponential smoothing method continues to remain poor.

Table 2.18: Accuracy measures for product code 814593 for daily sales

Model	RMSE	MAE	MAPE	MASE
Arima	*17.098*	*12.086*	*Inf*	*0.755*
NNE	*8.911*	*6.423*	*Inf*	*0.401*
ETS	*20.277*	*14.458*	*Inf*	*0.903*
HW	*19.226*	*11.834*	*Inf*	*0.787*

Next we experimented with product code *4610600*. Figure 2.27 shows how different models have fit the daily sales data for product code 4610600.

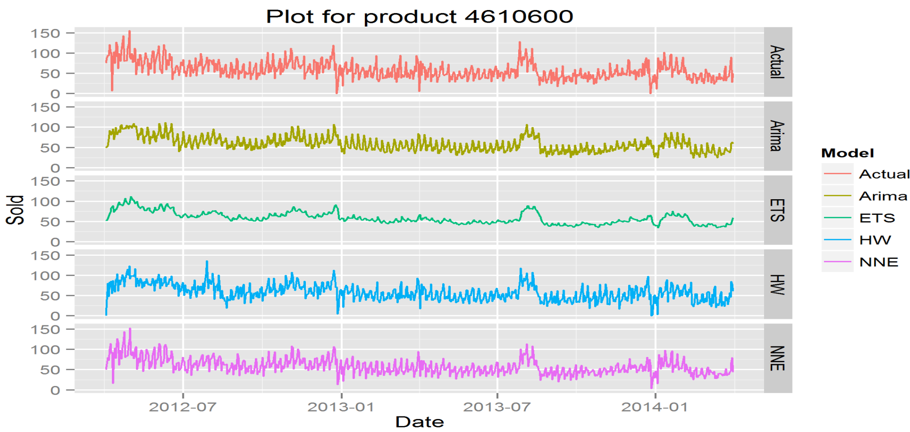

Figure 2.27: Data fit for product code 4610600 for daily sales

From our experimental results we saw that it uses *ARIMA(5,1,5)*, which means that it includes a non seasonal AR(5) term and a simple differencing of order 1 along with a non seasonal AR(5).

The arima coefficients observed from the results are presented in table 2.19.

Table 2.19: ARIMA coefficients for product code 4610600 for daily sales

	AR1	AR2	AR3	AR4	AR5
Coeff	*0.703*	*-1.360*	*0.658*	*-0.916*	*-0.097*
SE Coeff	*0.049*	*0.040*	*0.071*	*0.040*	*0.049*
	MA1	**MA2**	**MA3**	**MA4**	**MA5**
Coeff	*-1.416*	*1.851*	*-1.627*	*1.390*	*-0.601*
SE Coeff	*0.045*	*0.046*	*0.067*	*0.047*	*0.041*

The indicator values were seen as

AIC=11817.04 AICc=11817.23 BIC=11875.18

The ETS model selected was *ETS(A,N,N)* which means it has additive errors, no trends and no seasonality. This model can be written as

$$y_t = l_{t-1} + \varepsilon_t$$
$$l_t = l_{t-1} + \alpha\varepsilon_t$$

where the parameters as seen in the results are $\alpha = 0.1919$ and initial values $l_0 = 89.4607$.

The indicator values shown were :

AIC=18969.20 AICc=18969.21 BIC=18979.77

The holtwinters makes use of an additive model and shows the following values for the smoothing parameters :

$$\alpha : 0.1549$$
$$\beta : 0$$
$$\gamma : 0.2294$$

The neural network model again makes use of an average of 20 networks, each of which is a 30-16-1 network with linear output units.

The figure 2.28 shows the forecasts of all the four models discussed above.

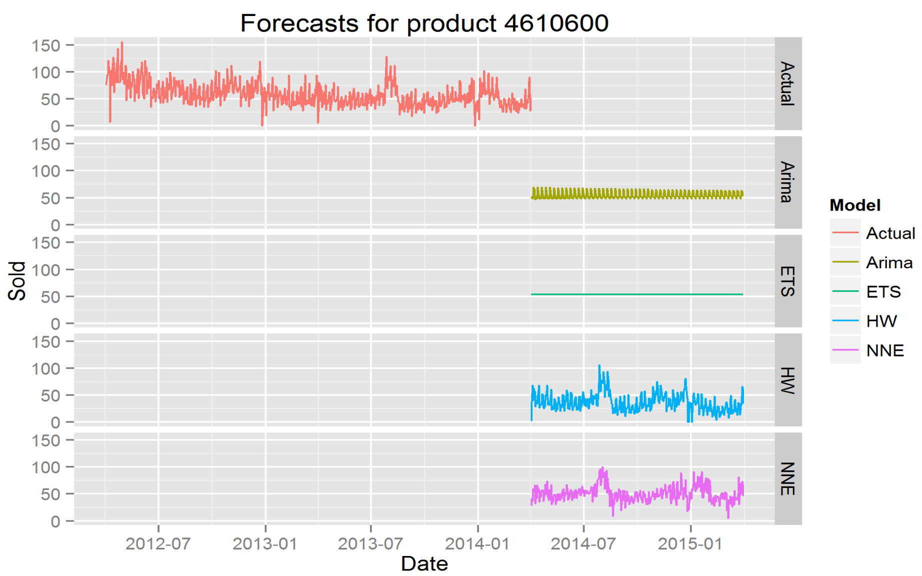

Figure 2.28: Forecasts for product code 4610600 for daily sales

We will now see how all the models performed by using the accuracy measures obtained from the results. The results are shown in table 2.20.

Table 2.20: Accuracy measures for product code 4610600 for daily sales

Model	RMSE	MAE	MAPE	MASE
Arima	*13.777*	*10.231*	*Inf*	*0.678*
NNE	*7.013*	*5.194*	*Inf*	*0.344*
ETS	*17.402*	*13.300*	*Inf*	*0.882*
HW	*14.843*	*9.138*	*Inf*	*0.585*

Neural network again outperforms the other methods and exponential smoothing still does not improve. We can easily say that exponential smoothing should not be preferred for daily sales sales forecasting.

We also performed some experiments to see how the models perform when only 2 years of data is provided instead of 4 years. To do so we made use of product code *4610600*.

Table 2.21: Accuracy measures for product code 4610600 for daily sales with 2 years of data

Model	RMSE	MAE	MAPE	MASE
Arima	*14.829*	*11.152*	*Inf*	*0.741*
NNE	*NA*	*NA*	*NA*	*NA*
ETS	*17.328*	*13.218*	*Inf*	*0.877*
HW	*NA*	*NA*	*NA*	*NA*

Table 2.22: Accuracy measures for product code 4610600 for weekly sales with 2 years of data

Model	RMSE	MAE	MAPE	MASE
Arima	*65.150*	*49.467*	*13.569*	*0.911*
NNE	*NA*	*NA*	*NA*	*NA*
ETS	*76.629*	*53.793*	*15.694*	*0.990*
HW	*47.200*	*27.240*	*6.549*	*0.494*

Table 2.23: Accuracy measures for product code 4610600 for monthly sales with 2 years of data

Model	RMSE	MAE	MAPE	MASE
Arima	*289.243*	*238.33*	*13.759*	*0.926*
NNE	*NA*	*NA*	*NA*	*NA*
ETS	*238.049*	*199.857*	*11.742*	*0.776*
HW	*264.049*	*175.702*	*8.867*	*0.708*

From table 2.21, 2.22 and 2.23 we can see that neural networks and holtwinters fail to give any solutions for daily sales, though holtwinters provides solution for lower granular data i.e weekly and monthly sales data, neural network continues to fail due to lack of enough data. ARIMA and exponential smoothing methods work fine with less amount of data as well. Also we found that ARIMA and exponential smoothing methods performed almost as good with 2 years of data as they did with 4 years of data.

Another interesting test that we performed was to check how the models behave while forecasting for products whose sales show linear relationship. For this purpose we selected product codes *4616753* and *4616760* which are shampoo and conditioner respectively, belonging to the same brand. The figure 2.29 shows that most of the times when a customer bought this brand of shampoo he also bought conditioner.

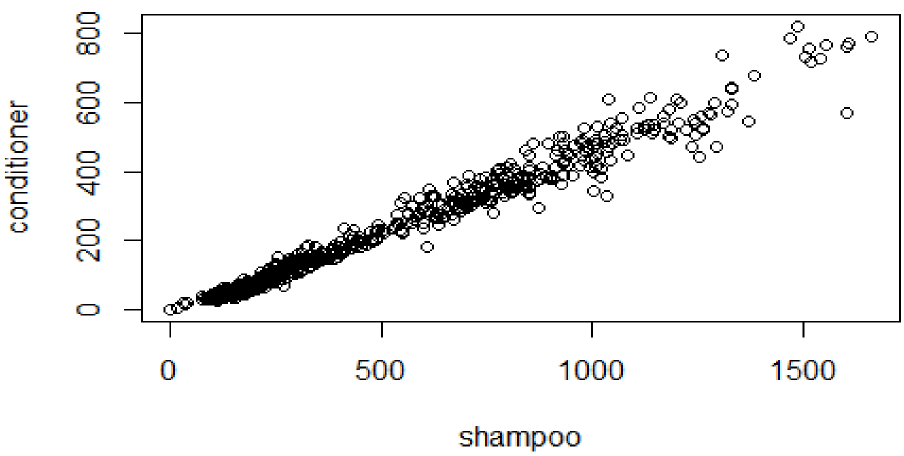

Figure 2.29: Scatterplot showing linear relationship between daily sales of products 4616753 and 4616760

We found that the forecasts for these products were very similar and so were the performance of the various forecasting techniques. This can be seen in tables 2.24 and 2.25.

Table 2.24: Accuracy measures for product code 4616753 for daily sales

Model	**RMSE**	**MAE**	**MAPE**	**MASE**
Arima	*153.355*	*98.771*	*Inf*	*0.941*
NNE	*87.226*	*50.596*	*Inf*	*0.482*
ETS	*168.838*	*104.280*	*Inf*	*0.994*
HW	*166.444*	*91.165*	*Inf*	*0.845*

Table 2.25: Accuracy measures for product code 4616760 for daily sales

Model	**RMSE**	**MAE**	**MAPE**	**MASE**
Arima	*72.625*	*47.763*	*Inf*	*1.001*
NNE	*49.068*	*27.796*	*Inf*	*0.582*
ETS	*78.082*	*47.184*	*Inf*	*0.989*
HW	*78.224*	*41.364*	*Inf*	*0.846*

We see that the MASE values are almost the same for both the product codes. Hence we can say that if we cluster all the SKU's which have linear relationship with each other i.e their sales patterns are very similar, then we can generalize and say that one particular method can be used and will provide the best results for all these SKU's.

Chapter 3

Conclusions

This chapter presents conclusions drawn from results that have been obtained from experiments performed on the data set and provide answers to the research questions. It also presents the limitations we faced and suggests how the results from this research can be improved in the future.

3.1 Conclusions

It has been observed that most of the companies which manufacture consumer packaged goods either make use of traditional techniques or the experience and intuition of their managers to predict future sales and determine the inventory levels of various SKU's. Since they have suffered from lack of precision in their forecasting methods, they are now looking at ways to improve their methods to have have comparatively accurate forecasts in order to minimize losses. With this as the motivation we set some goals at the beginning of this research and performed experiments to achieve them.

This research was performed in the form of a comparative study, where we compared traditional forecasting techniques with heuristic techniques to forecast sales based on historical sales data. Based on the literature survey we selected ARIMA, Exponential smoothing and Holt Winters, which are traditional forecasting techniques and compared their performance with artificial neural network model. We successfully performed modelling of these methods and wrote an automated script for prediction in R programming environment.

From our experiments we found that the feed forward neural network method performed the best at all levels of granularity, the only downside we found about this method was that it requires a lot of data to forecast accurately. When tried to forecast with 2 years worth of data, the neural network method failed to provide results, where as with 4 years worth of data its performance was way ahead of other traditional methods. It also handled seasonality, trends and irregular time series well.

We also found that the exponential smoothing method performed almost as good as the neural network method while forecasting monthly sales data, but when it came to short term forecasting i.e weekly and daily sales its performance was very poor and hence it is really not recommended for short term forecasting where as it is very suitable for long term forecasting.

ARIMA on the other hand performed well for short term forecasting where as it suffered while forecasting long term. The holt winters method seemed pretty constant and performed fairly well for both short and long term forecasts.

But in general it was seen that the performance of all traditional methods suffered badly in comparison to neural network method while forecasting on irregular time series.

Another observation we made was that it is not possible to say one particular method would work the best for a set of SKU's if they were clustered according to their product category or sub category, but on the other hand we found that SKU's which had similar sales pattern i.e their sales showed linear relationship could be clustered and the forecasting techniques showed almost the same behaviour for all these SKU's.

To summarize and to answer the research questions which we had set as per the concerns of the FMCG company whose data set we have used in this study, we can conclude as follows :

- As short term forecasting was their main concern, I would suggest them to use neural networks as they performed the best for all kind of patterns, but they must make sure that they have atleast 3 years of data. Otherwise they can use either ARIMA or holt winters, but exponential smoothing much be avoided.
- As far as data is concerned it would be better if they can collect as much historical data as possible, since the more the data points we have the better will be the accuracy of the forecasts. A minimum of 3 to 4 years of data is a must to atleast get some suitable results. Also I would suggest them to collect other information like promotions, price hike or reduction etc since these factors also will help us assess the data better.

Also though the neural network performs the best under all conditions, it is computationally difficult compared to other techniques discussed in this study. Also the results are equally difficult to interpret. Hence my advice to company would be that if they choose to use this method for their sales forecasting they should have an in house expert to help them in the process.

3.2 Research limitations

In general like most of the researches, even our study had some limitations which have been listed as follows:

- Since sales data is very confidential information, it was a herculean task to collect the data in the first place. Main task was to convince the managers at the company to provide us the necessary data, and then once we got the data we had to make sure that we select a good data set which will help us in our study as we found that a lot of data sets provided did not have enough data which is generally required for any forecasting based study.
- Due to lack of enough data points we could not forecast on yearly basis.
- We also did not have information regarding factors that could affect the sales, like promotions, discounts, price hike or reduction. These sudden changes might have an affect on the accuracy of the forecasts.
- We faced some limitation in the R neural networks package, due to the inability to visualize the results it was difficult to interpret them .

3.3 Future Research

As part of future work on this study, it would be interesting to see how hybrid models behave while forecasting using low and high granular sales data. Will their performance be better than the neural networks model? Also if information regarding promotions and other factors which drastically affect sales including competitors details like when they had promotions or launched a new product can be obtained, then even those details can be used to improve the accuracy of the sales forecasts.

Bibliography

[1] Koch, C. Nike Rebounds: How (and Why) Nike Recovered from Its Supply Chain Disaster [CIO magazine]. June 2004, [Cited 2014-04-10]. Available from: `http://www.cio.com/article/2439601/supply-chain-management/nike-rebounds--how--and-why--nike-recovered-from-its-supply-chain-disaster.html`

[2] Simatupang, T. M.; Sridharan, R. The collaborative supply chain. *International Journal of Logistics Management*, volume 13, no. 1, February 2002: pp. 15–30.

[3] Armstrong, J. S. *Long-Range Forecasting: From Crystal Ball to Computer*. New York: John Wiley and Sons, 1985.

[4] Moon, M. A.; Mentzer, J. T.; Smith, C. D. Conducting a sales forecasting audit. *International Journal of Forecasting*, volume 19, no. 1.

[5] ira Ktena, S.; Petropoulos, F.; Koutsoliakos, P.; et al. Forecasting Sales in a Sugar Factory. *Journal of Knowledge Management, Economics and Information Technology*, , no. 7, December 2011.

[6] Peterson, R. T. Forecasting practices in the retail industry. *Journal of Business Forecasting*, volume 12, 1993: pp. 11–14.

[7] H.Friedman, J. Multivariate adaptive regression splines. *The Annals of Statistics*, volume 19, no. 4, March 1991: pp. 1–141.

[8] Franses, P. H.; Draisma, G. Recognizing changing seasonal patterns using artificial neural networks. *Journal of Econometrics*, volume 81, no. 1, November 1997: p. 273–280.

[9] Alon, I.; Qi, M.; Sadowski, R. J. Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional methods. *Journal of Retailing and Consumer Services*, volume 8, 2001: pp. 147–156.

[10] Lee, W. I.; Shih, B.; Chen, C. Y. A hybrid artificial Intelligence sales-forecasting system in the convenience store industry. *Human Factors and Ergonomics in Manufacturing and Service Industries*, volume 22, no. 3, May/June 2011: pp. 188–196.

[11] Zhang, G. P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, volume 50, 2003: pp. 159–175.

[12] Zhang, G. P.; Patuwo, B. E.; Hu, M. Y. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, volume 14, 1998: pp. 35–62.

[13] Makridakis, S.; Wheelwright, S.; Hyndman, R. *Forecasting methods and applications*. New York: John Wiley and Sons, 1998.

[14] Dasgupta, C. G.; Dispensa, G. S.; Ghose, S. Comparing the predictive performance of a neural network model with some traditional market response models. *International Journal of Forecasting*, volume 10, no. 2, 1994: pp. 235–244.

[15] West, P. M.; Brockett, P. L.; Golden, L. A comparative analysis of neural networks and statistical methods for predicting consumer choice. *Marketing Science*, volume 16, no. 4, 1997: pp. 370–391.

[16] Fish, K. E.; Barnes, J. H.; Aiken, M. W. Artificial neural networks:a new methodology for industrial market segmentation. *Industrial Marketing Management*, volume 24, 1995: pp. 431–438.

[17] Ray, B.; Palmer, A.; Bejou, D. Using neural network analysis to evaluate buyer-seller relationships. *European Journal of Marketing*, volume 28, no. 10, 1994: pp. 32–48.

[18] Kang, S. Y. An investigation of the use of feedforward neural networks for forecasting. 1992.

[19] Gorr, W. L. Neural networks in forecasting: Special section Editorial Research prospective on neural network forecasting. *International Journal of Forecasting*, volume 10, 1994: pp. 1–4.

[20] Nelson, M.; Hill, T.; Remus, B.; et al. Can neural networks be applied to time series forecasting learn seasonal patterns: an empirical investigation. In *Proceedings of the 27 Annual Hawaii International Conference on System Sciences*, 1994, pp. 649–655.

[21] Faraway, J.; Chatfield, C. Time series forecasting with neural networks: a comparative study using the airline data. *Applied Statistics*, volume 47, 1998: pp. 231–250.

[22] Hill, T.; O'Connor, M.; Remus, W. Neural network models for time series forecasts. *Management Science*, volume 42, no. 7, July 1994: pp. 1082–1092.

[23] Foster, W. R.; Collopy, F.; Ungar, L. Neural network forecasting of short, noisy time series. *Computers and Chemical Engineering*, volume 16, no. 12, 1992: pp. 293–297.

[24] Chen, G. K. C.; Winters, P. Forecasting peak demand for an electric utility with a hybrid exponential model. *Management Science*, volume 12, 1966: pp. 531–537.

[25] Dungan, M. T.; Shriver, K. A.; Silhan, P. A. How to forecast income statement items for auditing purposes. *Journal of Business Forecasting*, volume 13, 1994: pp. 22–26.

[26] Alon, I. Forecasting aggregate retail sales: the Winters' model revisited. In *The 1997 Annual Proceedings.Midwest Decision Science Institute*, 1997, pp. 234–236.

[27] Sharda, R.; Patil, R. Neural networks as forecasting experts: an empirical test. In *Proceeding IJCNN Meeting*, volume 2, 1990, pp. 491–494.

[28] O'Donovan, T. M. *Short Term Forecasting: An Introduction to the Box-Jenkins Approach.* New York: John Wiley and Sons, 1983.

[29] FMCG Industry [ECONOMYWATCH Online]. June 2010, [Cited 2014-04-10]. Available from: `http://www.economywatch.com/world-industries/fmcg.html`

[30] Malhotra, S. A Study on Marketing Fast Moving Consumer Goods (FMCG). *International journal of innovative research and development*, volume 3, no. 1, January 2014.

[31] Majumdar, R. *Product Management in India.* New Delhi: Prentice-Hall of India Private Limited, 2007.

[32] Hines, P. The Principles of the Lean Business System [Online]. 2010, [Cited 2014-04-10]. Available from: `http://www.atem.org.au/uploads/publications/-The_Principles_of_The_Lean_Business_System.pdf`

[33] Jina, J.; Bhattacharya, A. K.; Walton, A. D. Applying lean principles for high product variety and low volumes: some issues and propositions. *Logistics Information Management*, volume 10, no. 1, 1997: pp. 5–13.

[34] Helo, P. Fast moving consumer good - a productivity perspective on supply chains. *International journal of Productivity and Quality Management*, volume 5, no. 3, 2010: pp. 269–295.

[35] Hyndman, R. J.; Athanasopoulos, G. *Forecasting: principles and practice.* OTexts, 2013. Available from: https://www.otexts.org/fpp/

[36] Hyndman, R. J.; Koehler, A. B. Another look at measures of forecast accuracy. *International journal of forecasting*, volume 22, no. 4, 2006: pp. 679–688.

[37] Hyndman, R. J.; Khandakar, Y. Automatic time series for forecasting: the forecast package for R. Technical report, Monash University, Department of Econometrics and Business Statistics, 2007.

Appendix A

Acronyms

FMCG Fast Moving Consumer Goods

ARIMA Autoregressive Integrated Moving Average

ANN Artificial Neural Networks

SKU Stock Keeping Unit

CPG Consumer Packaged Goods

DC Distribution Centers

POS Point of Sale

VMI Vendor Managed Inventory

AR Auto-Regressive

MA Moving Averages

RMSE Root Mean Square Error

MAPE Mean Absolute Percentage Error

MASE Mean Absolute Scaled Error

MAE Mean Absolute Error

ACF Auto-Correlation

PACF Partial Auto-Correlation

MLP Multi Layer Perceptron

SAR Seasonal Auto-Regressive

AIC Akaike Information Criterion

AICc AIC with correction

BIC Bayesian Information Criterion

Appendix **B**

Contents of CD

- `readme.txt` .... the file with CD contents description
- `data` .... the data files directory
  - `graphs` .... the directory of graphs of experiments
    - `daily plots` .... the directory of graphs for daily sales
    - `weekly plots` ....the directory of graphs for weekly sales
    - `monthly plots` ....the directory of graphs for monthly sales
- `src` ....the directory of source codes
  - `code` .... the directory of R source code
    - *model_script_new.R*......................the automatic R script
    - *datapoints_new.csv* .... the data set
  - `thesis` ....the directory of LaTeX source codes of the thesis
    - `figures` .... the thesis figures directory
    - *DP_Ravichandran_Ruhi_*2015`.tex` . . the LaTeX source code files of the thesis
- `text` ....the thesis text directory
  - *DP_Ravichandran_Ruhi_*2015`.pdf` the Diploma thesis in PDF format
  - *DP_Ravichandran_Ruhi_*2015`.ps` . . . the Diploma thesis in PS format