Adaptive Forecasting

A survey of mathematical methods and modern learning techniques
in Inventory Management

Geert Roseboom
BMI-paper
April 2005

Vrije Universiteit Amsterdam
Faculty of Exact Sciences
Business Mathematics and Informatics
De Boelelaan 1081a
1081 HV Amsterdam

Preface

In the last year of the doctoral program Business Mathematics and Informatics (BMI) at the Vrije Universiteit in Amsterdam every student has to write a paper on a self-chosen subject related to its curriculum. This is the so-called BMI-paper. Because the aim of the BMI program is to combine the fields of Business Management, Mathematics and Computer Science, a BMI-paper should cover at least two of those three areas.

Since my bachelor internship at TNO-Physics and Electronics Laboratory (TNO-FEL) in The Hague (NL), where I worked on an implementation of the TABU-search heuristic for an optimization problem, I developed an interest for modern learning techniques, also known as computational or artificial intelligence. Even though these modern techniques sometimes lack solid mathematical background, they have been applied successfully to a wide range of practical problems. The result of this interest was that I chose my academic courses mainly in the field of artificial intelligence, programming and statistics. During the last year I also became interested in stochastic modeling and simulation.

After I followed his course on Data Mining I asked Wojtek Kowalczyk for an interesting subject for my BMI-paper, which would contain/combine elements like: simulation, modern learning techniques, forecasting and/or optimization. This resulted in the idea to apply these three approaches to a specific field of forecasting problems and investigate, which differences exist between modern learning techniques and conventional mathematical methods. Due to time limitations we restricted our research to a short literature study and some experiments.

At last I want to express a few words of gratitude. I would like to thank Sandjai Bhulai for his help finding the right mathematical articles and his patient, enthusiastic attitude whenever I came to him for questions. I thank Menno Dobber for putting one of his datasets at disposal. And finally I want to thank my supervisor Wojtek Kowalczyk for his support in my last academic year in general. He was always there and the interesting discussions with him made me feel aware of the rightness of my choice for this academic program.

I hope the reader will enjoy this paper.

Geert Roseboom

Wageningen, April 2005

Management summary

This paper aims to provide an insight in the role intelligent forecasting methods can play in the world of Inventory Management. An important assumption throughout this investigation is that data originating from real life processes can behave in a capricious and non-stationary manner. In such case adaptive methods, capable of making accurate forecasts while they need to respond to changes in demand over time, can bring relief and are a must to maintain control over the supply chain.

In the first part the fields are explored where sophisticated forecasting strategies can play a significant role. We pay attention to the difficulties and limitations that have to be overcome. To emphasize the surplus value these strategies can have for business we illustrate our story with some examples of recent implementations from major players in business intelligence such as SAS, SAP and IBM. Clearly they see interesting opportunities too, as they have set in a trend with the development of a wide range of applications that are all presented as sophisticated demand forecasting algorithms.

Unfortunately these implementations are more an exception than a rule. Therefore one of the goals of this paper is to contribute to the publicity of these intelligent forecasting techniques and make the reader aware of the opportunities for businesses.

The second part of this paper can be viewed as an introduction to the methods that lay on the foundation of these applications. The conventional techniques to solve forecasting problems have their roots in applied mathematics. However, in the last decade modern learning techniques, such as artificial neural networks, have caught up fast and their successful implementations are encouraging to build new applications. Since companies rarely give away competitive information, we turned to literature and looked for mathematical as well as modern learning methods. Subsequently we selected the most common approaches, which are:

- Bayesian
- Stochastic Approximation
- Artificial Neural Networks

In the currently available literature it remains unclear which of these approaches deserves recommendation. We found just one article that made an actual comparison between mathematical and artificial intelligence techniques. An explanation for this lack of comparisons is that, traditionally these two schools approached problems from different perspectives. During this investigation we learned that the typical contrasting aspects between the schools are diminishing, so it is likely to expect that comparisons will be made in the near future.

The lack of comparisons made us feel challenged to setup our own very basic experiments with interesting real life datasets. The primary focus was to develop a feeling for the performance of our selection of methods. In these experiments the neural networks were the clear winner. Although the results are of modest importance from a scientific point of view; it provides a basis and hopefully encourages people to do further work in comparing mathematical approaches with modern learning techniques.

Content

1. Introduction

In this paper we shall explore the world of forecasting and the value forecasting methods have in optimizing business objectives. Although this is tremendously interesting, with the huge amount of available information sources in our information era it is also quite ambitious. To narrow the scope of this project and give it a concrete bias, we have chosen to put our focus on a specific area of application: Inventory Management (IM).

In the first paragraph of this chapter we give a short description of the field of IM. The second paragraph contains the problem description/research objective. In paragraph three we describe the research methodology. The last paragraph gives an overview of the contents of each chapter in this report.

1.1 Context

In the last decades manufacturers and companies have moved from a production push environment, which is merely focused on achieving production efficiency in single chains of the total logistic process, to a consumer pull environment, that is mainly focused on meeting the expected consumer demand.

Therefore controlling and fully understanding the consumer demand is crucial to have an effective grip on product costs and customer satisfaction. This control phenomenon, also known as Demand Management (DM), is considered as the most important factor in Supply Chain Management (SCM).

The objective of SCM is to bring the penetration point as close to the supply as possible (Figure 1.1 gives an overview of the activities within SCM). Since in most cases product demand is not known in advance, it is necessary to make forecasts/predictions of the demand. Seen from this perspective, looking after a proper DM becomes a key driver to the success of any SCM enterprise.

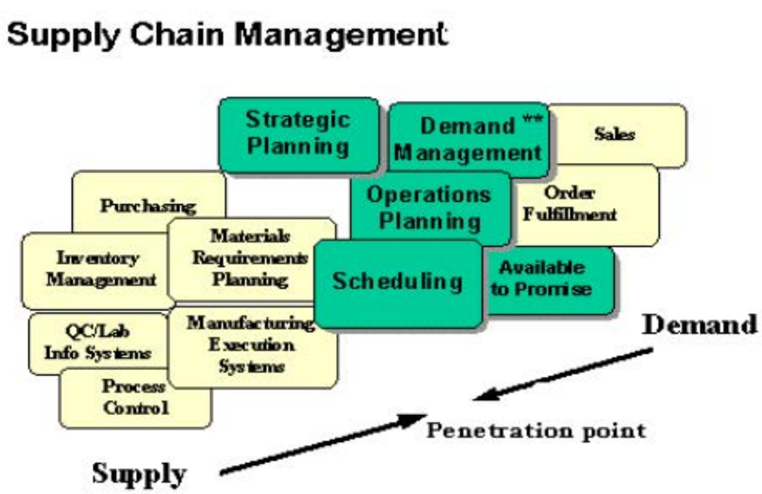

Figure 1.1 Activities within SCM applications [22]

In practice, frameworks that support or make possible SCM are developed by leading software companies like IBM, Microsoft, Oracle, SAS, and SAP. At the moment these companies are taking a step further as they are starting to concentrate on achieving higher levels in SCM. For example by implementing tools that can accurately and autonomously predict sudden increases in workload demand.

The application environment for forecasting methods in IM can be seen from a broad perspective; for example also in private use we can expect that the future will bring us several personal forecasting applications; e.g. imagine a refrigerator that keeps track of a person's groceries and automatically reorders the products whenever its stock is below some threshold.

Of course such ordering systems can be applied to many other processes, and each ordering system for each process will be implemented differently. The theoretical field that deals with the development of methods/strategies that lay on the ground of these ordering systems is known as Inventory Management. With these methods, which can be mathematical approaches and/or modern learning techniques, it is the goal to forecast demand, gain control and optimize as many aspects of the process at hand.

1.2 Problem description

For centuries mankind has endeavored to forecast all kinds of phenomena out of a variety of interests. The idea and desire to forecast processes can be applied to almost any process we put our mind at. In the past this concerned mostly natural phenomena such as predicting the orbit of a planet. Nowadays forecasting methods are used as a way to control and optimize business objectives. This can be a considerably more demanding process, especially if the underlying processes are not behaving according to some stationary pattern.

Whenever working with real-life data such non-stationary behavior will be a recurring and standard situation. In most practical cases the data will be difficult to observe, and conclusions one can draw will be based on partial information. Fortunately with time forecasting methods have evolved, which should enable us to make predictions even when data changes over time. The field that tries to make such sophisticated predictions is christened Adaptive Forecasting.

With this introduction we arrive at the main objective of this paper, which is to provide an insight/ overview in the role Adaptive Forecasting methodologies can play in the field of Inventory Management. To gain this insight we divide the main objective in the following phases:

- General introduction to illustrate recent developments within IM/SCM
- Review of selected mathematical and artificial intelligence forecasting procedures
- Experiments with these procedures on real-life datasets

The general introduction is meant to make the reader familiar with the most important aspects and methods in Inventory Management. We briefly discuss how the field has developed and what the requirements on forecasting methods are today. Rather than to get lost in details, we just discuss general issues and some of the advantages and disadvantages of available techniques. Next to discussing the features that should be captured by intelligent forecasting procedures we look at some recent business implementations.

In the second part of this investigation we have made a selection from techniques that form the heart of these business implementations. In our selection we have looked to what extend our methods are capable of taking into account properties like: adaptability to deal with the changing customer behavior over time, ability/necessity to incorporate historical data in the model. In the report we attempt to remain concrete and illustrate the results in the field as much as possible by implementing an example of a stock-ordering model for a fictitious warehouse. A warehouse can be seen as a shop that offers a collection of products. We will focus on an ordering system for a single product, which will be defined in more mathematical terms later in this report.

Each of such ordering models will have to be configured to satisfy the wishes of its user. This means that the ordering system will have to deal with the arrival pattern/demand of the customers, which is responsible for the amount of goods taken from the shelves. For this reason we can see the arrival pattern as a direct consequence of human behavior. As we all know, human behavior is complex and its changes over time will have consequences for the demand.

The data that results from observing such complex behavioral patterns forms an excellent opportunity to find out which forecasting methods are more suited to deal with real-life data, modern learning techniques or the more conventional mathematical approaches. Therefore we have included a few experiments with some of the reviewed techniques.

1.3 Research Methodology

The resources that were consulted to find the articles and background information for this study can be divided in three categories:

- Field expert knowledge; from:
  - Mathematical articles and general advice from Dr. S. Bhulai
  - Artificial Intelligence articles and general advice from Dr. W. Kowalczyk
- Library search; in the contents of (the last five years) magazines as:
  - Management Science
  - Operations Research
  - Annals of Mathematics and Artificial Intelligence
  - Artificial Intelligence in Engineering
  - Artificial Intelligence
  - Machine Learning
- Internet search; using search engine Google with names, keywords (or combinations of keywords) like:
  - {Forecasting, Adaptive, Inventory, Management, Survey, Control}
  - Names from authors encountered on previous articles
  - Sites from companies that are likely to offer adaptive forecasting services

With the gained knowledge we made a simple experimental setup to test (some of) the collected methods on real life data sources.

1.4 Structure of the report

The report is built up as following; Chapter 2 gives a brief survey of developments in the area of our interest over the last two decades. This is done to elucidate the role intelligent demand forecast procedures can play for business objectives. Chapter 3 contains the description of our example shop model. Armed with this model we proceed to discuss our selection of mathematical and artificial intelligence techniques. The simple experiments performed to develop a feeling for some commonly used forecasting techniques can be found in Chapter 4 together with its results. In Chapter 5 the conclusions are presented and discussed.

2. Intelligence in Business Applications

Before discussing which forecasting techniques are available in Inventory Management, let us become acquainted with developments in supply chain and demand management over the last decades and emphasize the role intelligent demand forecast procedures can play for business objectives. In the second paragraph we will go into details with some actual implementations/frameworks from the major companies that operate in the field of our interest.

2.1 Developments in Supply Chain and Demand Management

In the early years of the twentieth century, industrialization made it possible to efficiently and effectively produce on a large scale against relatively low costs. Since companies are always under influence of a competing market, there exists a constant pressure to improve their processes in such a manner that they can meet demand and stay ahead of competition.

The next major contribution to this efficiency strike has been the development of computers. Due to computers we have become able to automate and organize processes more efficiently. Simultaneously analytical methods have been developed to optimize production in a single link of the supply chain. In the latter decades one became aware that efficiency in a single link not necessarily leads to efficiency in the total chain. With this understanding we have caught the heart of what Supply Chain Management is all about; striving for efficiency in the total supply chain.

Pursuing an effective SCM goes hand in hand with the development of the appropriate analytical methods. From figure 1.1 we have learnt that all processes in SCM heavily depend on effective Demand Management. Given this significance, prudence is in order with the use of forecasting methods, since inaccurate product demand forecasts can easily ruin all the effort that is put in a carefully designed supply chain system.

A well-known effect that might have catastrophic consequences when using imperfect forecasting techniques in SCM is known as the bullwhip effect, which implies that demand variability increases as one moves up the supply chain i.e. as one moves away from customer demand (more details on this effect can be found in [5]). Consequences of such phenomena are expressed in excessive inventory levels, which can cause a company to end up with too much of its assets tied up in inventory, restricting its ability to respond to changes in its market environment etc. To prevent the occurrence of such effects man is constantly looking for the most accurate forecasting technique.

2.1.1 Demand Forecasting

Now we have become familiar with the importance of adequate demand forecasting, we take a look at their basic requirements. Of course such methods are not completely new when it comes to supply chain management applications. All manufacturers and marketers have some methodology of demand forecasting. The first attempts to automate this highly human dependent sector of the supply chain, were statistical models that predict future demand based on historical sales volumes. Usually these models were based on a single variable and rarely effects of other variables were incorporated in the demand model. Thanks to the rapid development of Personal Computers, with time it became possible to use multiple regression and other similar statistical techniques on a large scale which opened doors for more accurate prediction methods.

Despite these possibilities, we learn from literature that most available models still ignored variables that influence demand. In most cases the ignoring is based on the assumption that these variables would not change significantly. One argument to brush aside such variables was that they were subjective by themselves and did not lead to better forecasts. The common opinion was that the best way to account for the effect of their changes is to combine the generated forecast with an expert human opinion, for more details see [10, 11, 23]. Although we live in a highly automated society this does not seem as a very strange opinion yet. At the moment, still a lot of research is conducted to develop methods that are capable to weigh in expert opinions.

As we look at social developments, the recent trend is that 24-hour economies and changes in our modern society, such as individualism, secularization etc., have increased diversity in the consumption patterns of populations. With such an increase we can imagine that the amount of relevant parameters increases too and make simplified models inadequate. Simplifications can defeat the very purpose of building these models when outcomes become so generic that they fail to explain the specific behavior under investigation.

Preceding arguments have shown us that demand forecasting can be viewed from many different perspectives. This brings us to the following reasoning that tries to explain why it is so crucial to dispose over the appropriate forecasting techniques; the more perspectives involved the more complex the situation, the more complex the problem the greater the reward and the more money involved the greater the need and expectations on demand forecasting methods. After all in our case, the player that is capable of effectively organizing its inventory is most likely to end up as the player with the best competitive position.

2.1.2 Difficulties with conventional Demand Management methods

By tradition, the answer to the need for appropriate demand forecasting procedures was found in statistical non-linear regression algorithms. These techniques involve fitting functions into the historical demand function and then using the developed functions to predict future demand. Such techniques are straightforward in its use because they have their roots in applied mathematics and will therefore be familiar for most technically skilled users. The resulting out-of-sample forecasts were claimed to allow extrapolation beyond historical operation regions, but this only holds true when the functions have not been tuned on a specific region.

A basic problem with function fitting is that the user is required to know which type of function will best describe the demand. Even for skilled professionals it remains difficult to decide if the equations will satisfactorily fit the complex trends of consumer demand. In most cases models will be based on a group of functions, where each function is used to describe a different region of the demand function. And a great potential danger of the out-of-sample forecasting lies in the fact that new developments might behave in a completely different way.

Other problems arise when mathematical relationships are difficult to expose, also for experts it can be quite hard to discern the key variables that affect demand. Whether the reason behind it is lack of knowledge or just a very complex process, it often requires that simplifying assumptions have to be made to make the problem tractable. Even when these simplifying assumptions can be made, statistical modeling requires a significant amount of expert effort to develop adequate models. It can be extremely disappointing whenever it turns out that the models not function as well as hoped due to unavoidable whimsicalities of real life processes. Therefore one is constantly looking for methods that generalize well and yet remain accurate.

2.1.2 Expectations on future Demand Management methods

As we have stated before, today's markets are in constant movement as they need to respond to a variety of changing influences like:

- Seasonality
- Product life cycle
- Economic developments
- Etc

Products with short life cycles, for example mobile telecommunication devices, need fast and accurate predictions of the amount of product that is needed. In an ideal situation models are forehand even before going to market, so that reservations made by enthusiasts prior to an introduction can help forecast actual demand over the product life cycle.

Next to that the models should be able to easily process the extra information that comes available during their lifecycle, as to continuously update predictions. With the introduction of new technologies like Radio Frequency Identification (RFID) Tags (more technical details can be found in [1]) more accurate demand and sales data become available, which offers excellent tracking opportunities to respond to customer behavior.

The shortcomings of conventional methods open the field of forecasting for new technologies that can support marketing and sales in decision making and actually take into consideration all variables which can influence demand for a given product. From the mathematical corner advances are made in the field of Bayesian methods, stochastic approximation algorithms and models based on Markov Decision problems. Eyes are also fixed upon modern approaches like artificial neural networks and reinforcement learning, which are successfully applied to a widening scale of problems. The idea is that these generic, black box models built on historical sales data are the answer to the need for robust and accurate demand forecasting models.

2.2 Business Intelligence: Implementations

In the following sections we attempt to take away some of the mystic/profoundness of business intelligence methods by providing some illustrative examples of recent implementations in this field. Three examples cover summaries of success stories taken from the well-known business intelligence companies' websites like SAS, Cap Gemini and IBM. The last example is that of Evant, a company that develops demand planning software for the retail industry, whose encouraging story shows that advancements in this field are not solely made by the big companies.

2.2.1 IBM

From the internet article "IBM unveils 'self-healing' tools" by Ann Bednarz of 03/10/2003 on the IBM website, we can conclude that also IBM has acknowledged the surplus value of intelligent forecasting methods. With the introduction of a new dynamic provisioning tool IBM claims to have moved closer to achieving its vision of self-healing, self-configuring systems. The tool is designed to predict and respond to sudden increases in data center workloads.

The actual software consists of three modules: The Adaptive Forecasting module, which uses mathematical models to anticipate to the progression of an unexpected surge in demand. The Online Capacity Planning module, that estimates the resources required to maintain service-level targets during peaks and allows a hot swap of resources from one workload to another. And the Rapid Reconfiguration piece, which uses new capabilities in WebSphere Application Server 5.0 to add and remove nodes as resource demands fluctuate.

In the past IBM already had forecasting tools that can prepare systems for predictable spikes, such as seasonal traffic bursts. But dealing with unexpected surges requires something more sophisticated. When workload starts to deviate significantly from what is expected, IBM's new tool will start to track how things are changing. By using probabilistic methods that can deal with short-term forecasts the system tries to get ahead of the surge, and makes the necessary requests for the appropriate amount of server power to prevent service levels from degrading. Of course, when the surge decreases, the tool can release resources to conserve costs.

2.2.2 SAS-CapGemini

One of the most recent examples of an adaptive forecasting implementation is literally named "Adaptive Forecasting" and has been introduced on June 7, 2004 by SAS & CapGemini. As SAS is considered the world leader in business intelligence, CapGemini is well-known as one of the world's largest providers of consultancy, technology and outsourcing services. Together they have launched this application what has to be an "Innovative Forecasting Solution for Utility Industry".

The Adaptive Forecasting application is initiated by a 2003 benchmarking study by CapGemini. The study of ten US utility companies (also known as the sector of energy industry) shows that the ability to more accurately predict the volume of supply and demand along with more accurate revenue projections can result in significant profits for those who do it well. The study concludes that many utility companies find it difficult or impossible to achieve these savings due to a lack of integration, flexibility, and functionality in their current demand and revenue forecasting systems. The newly developed application tackles these problems as it aims to provide "great breadth and depth of modeling techniques and an unsurpassed level of forecasting accuracy, flexibility, and scalability in handling large volumes of data which previously hindered the energy industry."

2.2.3 SAS – Implementation for Kirin Brewery Company

Another success story from SAS is found in the Internet Article "Always Enough, Never Too Much", which concerns an SCM application that SAS developed exclusively for Kirin Brewery Company in Japan.

The Kirin spirits and wine business offers approximately 700 different alcoholic products and largely depends on imported products. They trade with thirty-four suppliers in fourteen countries. The lead time depends on several factors such as: the area, season, order lot, as well as the tonnage capacity of trucks, which is restricted by law in some nations. In the past, these processes were dependent on individual staff knowledge and experience and carried out with spreadsheets.

In April 2003 started the "Spirits and Wine SCM System" project, based on SAS Supply Chain Intelligence. As made clear before, failure to control inventory can have huge consequences on the whole business, i.e. lost sales opportunities caused by short stock and unnecessary costs incurred by excess inventory. The goal of this project was to improve the current supply/demand planning system to cope with such inventory issues. This resulted in a new system that is claimed to have brought "dramatic improvements in accuracy of demand forecast and automation of complicated procurement processes".

The new Spirits and Wine SCM System relies its success on two major strengths. One is that the procurement operation, including order planning and future stock simulation, previously carried out by individual staff, is systemized and automated. The other is the seven-model system specially developed for the spirits and wine business to forecast demand and prepare the shipping plan. The models are specially designed by arranging dozens of analytical methods, including methods based on past data, methods reflecting the latest performance, and methods for seasonality. Simulations are run using models with different variables and patterns chosen for product characteristics, subsequently users can choose the most relevant forecast among the seven.

Other issues, such as inabilities to incorporate forecast data, have also been taken care of in the new system. With this function, it has become possible to continuously monitor actual data on sales performance and other relevant figures. Furthermore, an alarm function is adapted to notify the user of deviations from the forecast in the actual figures, which enables them to take appropriate measures against the change.

The article concludes that the new system has brought many immediately noticeable positive results. One of the most significant results is that, instead of being attained with doing forecasts to determine the appropriate order volume themselves, staff employees can focus on verification of forecasted values and further sophistications of the newly used system.

2.2.4 Evant

The story of Evant should appeal to any enterprising academic person, given that it is founded to put theory into practice. Evant is co-founded by Professor H. Lee, an authority on supply chain and inventory management as he is: Professor of Operations, Information, and Technology at Stanford Graduate School of Business, founder and director of the influential Stanford Global Supply Chain Management Forum and current editor-in-chief of Management Science. In [15] Lee puts forward two forces that have great potential for intelligent, demand-based SCM.
First, companies with strong partnerships and sound information systems to track data flows across the supply chain can now turn those into intelligence. Second, companies that have streamlined their supply chains and improved their operating performance can realize bigger value by combining their marketing efforts and demand management with their supply-chain management initiatives.

These ideas are encapsulated in Evant`s technology, which has been successfully implemented for the pharmaceutical company Longs Drug Stores. Longs Drug Stores forms a chain with 470 retail outlets, good for an annual turnover of $4.5 billion. From their start in the late 30`s the company has emphasized excellent customer service. This was translated into a corporate strategy of never-empty shelves, which often comes hand in hand with lots of excess inventory.

Even though pharmacists and buyers were aware of the seasonal demand patterns for their major drugs, such as flu and allergy seasons, they were unable to master the complex inventory challenge. With Evant`s software, Longs established a next-generation inventory-planning tool. Since 1997 Longs' inventory has decreased 65% at its distribution centers, while the corresponding store inventory has dropped 38%. The company's annual savings from using scientific methods to run its data-rich demand chain was $36 million.

The system's methodologies should allow more accurate forecasts as it claims to optimize demand-chain activities as: forecasting, inventory control, transportation, material handling, and warehousing. Important is that the system is data-driven, which should minimize guesswork for most of the products. The core features of Evant`s technology are:

- Statistical analyses of day-of-week, seasonality, and trend effects on customer demand of products.
- Tracking and analysis of forecast errors to continuously fine-tune and self-adjust the demand-forecasting methodology.
- Explicit treatment of supply uncertainties, e.g. supplier fill rate and delivery performance.
- Integration of demand forecasting, inventory planning, and transportation logistics into planning decisions.

3. Forecasting models in Inventory Management

The previous chapters have given us a feeling for the strong influences demand forecasting can have on optimizing business objectives. For businesses this usually comes down to maximizing customer satisfaction while at the same time minimizing their inventory levels. So far our look at forecasting has been from a predominantly managerial perspective as we looked at examples of software modules that provide the forecasting environment. Unfortunately for us, companies rarely give away the techniques they have used to build their modules. Therefore we have consulted literature and other relevant sources to gather information on this particular topic.

For this paper we are interested in those techniques that are used to solve control and forecasting problems. In case of simple problems there is already a huge collection of conventional approaches that offer assistance in deciding which inventory levels should give an optimal solution. However, when there is not enough information available on the demand distribution and this distribution is not stationary it is uncertain whether these methods are also adequate to deal with these time varying conditions. Stationarity means that the main moments (mean, variance) of a distribution (or sample) do not change over time (i.e. mean and variance over a part of the sample are independent of where in the sample this part is extracted). In this chapter we take a closer look at methods from mathematics and artificial intelligence that should be more capable of tackling such problems.

In the first paragraph we start with an example of an inventory model for a little shop, which will function as an illustrative model for the mathematical techniques. Paragraph two contains different mathematical approaches that are available to solve the problems of a shop manager facing an unknown and/or non-stationary demand. Next to describing these policies in general and making them applicable to the example model, there is looked after to which extent they will meet needs like: adaptiveness to deal with changing customer behavior over time and the ability to incorporate historical data in the model. In paragraph three we treat the techniques artificial intelligence has to offer for this type of problems.

3.1 Problem formulation

Let us assume that we are put in the role of a shop manager of a very simple shop, which has exactly one product in its assortment. The job of the manager is to make as much profit as possible. This means that we have to find a strategy that can balance high ordering costs and product availability. We assume that our demand is unknown and that it can change over time. Hence we do not expect a simple order strategy. In this paragraph we define our illustrative model in a more mathematical form.

So we have become the manager of a shop with just one type of product and we have customers that are responsible for creating a certain demand for each discrete step in time. Suppose time is also discrete and we find ourselves at time unit t. At this moment we have a stock of x items and we are interested in the profit $V_t(x)$ we will have from now on.

At the beginning of each point in time we can order an amount of items. We will call this decision a. The magazine of our shop has a maximum capacity of m items, which means that we can maximally order $m-x$ items.

We assume that our order is delivered instantaneously against certain order costs that are given by the function $b(a)$. The order results in a new stock level of $x+a$ items, which are held in store for the entire period against costs h.

In that same period we experience a total demand for k items, which is given by probability $\pi_\theta(k)$. We assume that our unknown demand, which is generated by all our customers, can be modeled as a Geometric distribution with a certain unknown parameter θ. This means that for each step in time we have probability function $\pi_\theta(k)=\theta(1-\theta)^{k-1}$ that gives the demand probability for k items, with $k=1,2,\ldots$.

At the end of each period we have sold a certain amount of items against price w. This amount is the minimum of our available stock and the total demand of our customers in that period, which can be written as $\min(x+a,k)$. Note that this means that keeping inventory levels too low will result in lost sales and perhaps even more important for our estimate that we are unable to observe the true demand.

At time $t+1$ we will start with stock $(x+a-k)^+$, where the notation $(z)^+ = \max\{z,0\}$ is used to indicate that only non-negative values are allowed. Therefore in the next period our costs will be $\alpha V_{t+1}((x+a-k)^+)$, where $0<\alpha<1$ is used as a discount factor to represent the value depreciation between the current and the next state.

Because this problem returns for each point in time, we can write all our previous modeling work in one dynamic programming recursion step:

$$V_t(x) = \max_{a\in[0,\ldots,m-x]} \{\sum_{k=0}^{\infty} \theta(1-\theta)^{k-1}[\min(x+a,k)w - b(a) - (x+a)h + \alpha V_{t+1}((x+a-k)^+)]\}$$

with:

θ = *unknown parameter of demand probability*

x = *number of available items*

a = *order quantity*

k = *demand of customers*

m = *storage capacity*

w = *price of each item*

$b(a) = order\ \ \cos t$

$h = holding\ \ \cos t$

$\alpha = discount\ \ factor$

In our explanation we have assumed that we experience a demand for exactly k items, but this is unknown. Therefore in the recursion step we weight over all possible k with θ. Because we are speaking in terms of profit we take the maximum of all possible a `s we can choose, which is denoted as $\max\limits_{a \in [0,\ldots,m-x]}$.

Now we have specified a model that helps us to maintain an optimal level of our costs, we have to find techniques that provide us with accurate estimates for our unknown parameter θ of the demand distribution.

From the previous chapter we have learnt that conventional mathematical techniques are not suited to deliver adequate estimates when the demand function is unknown and changes over time. Note that for clearness this last property causes us to change θ in θ_t to express the time dependence, which will be the notation from this moment on.
In the next paragraphs we look into methods that should be capable of providing better estimates as more information becomes available.

3.2 Mathematical methods

In 1985 Kumar [13] produced a survey of results in stochastic adaptive control. He divides the field of Adaptive Control into two parts, Bayesian Adaptive Control Problems (BACP) and Non-Bayesian Adaptive Control Problems (NACP). Examples of NACP (in this article) are methods like self-tuning regulators and Markov chains. Since the paper aims to provide more insight in the techniques that were available in the total field of adaptive control, it does not refer to specific achievements concerning IM. The overall conclusions, at that time, were that still a lot of work was needed to provide efficient schemes for implementation of the algorithms and studies for the rates of convergence.

Following the by Kumar proposed division it comes quite naturally that in the next sections we shall discuss the concept, advantages and disadvantages of a Bayesian and a Non-Bayesian implementation, in this case Stochastic Approximation. The last section of this paragraph is devoted to a discussion of one of the few and recent articles on Adaptive Inventory Control.

3.2.1 Bayesian Adaptive Control

One of the earliest works on non-stationary (or varying) stochastic demand problems is ascribed to Karlin [12] and stems from 1960. Karlin analyzes a dynamic inventory model in which the demand distributions can change from one period to another and discusses concepts of partial information.

In the same period (1959-1960) H.E. Scarf was one of the first to work on Bayesian methods in adaptive control, which can be used as a method to solve stationary partial information problems. Until that time most dynamic inventory models with stochastic demand assumed that the underlying demand distribution was known with certainty. When using a Bayesian approach to estimate the unknown distribution function, the idea is that with each step in time more information will have been gathered about the actual distribution function of the demand process. Since with each update information is added, the shape of the distribution function has a sharper and more accurate variance.

In the beginning of the 80`s, Scarf and Karlin's work on inventory control and unknown demand was extended by Azoury [2, 3]. The Bayesian dynamic inventory problem can be modeled as a dynamic program with a multidimensional state variable. The state space is expanded to include one or more variables that describe past information about the demand distribution. This multidimensionality was considered as a serious disadvantage and has been an obstacle to apply Bayesian methods on inventory problems, since it brings along a greater computational complexity. In that time, it was often preferred to solve a non-Bayesian approximation to the Bayesian dynamic program. However, in [2] Azoury presented results that make it possible to determine the optimal Bayesian policy by the recursive computation of functions of one variable only, which makes it more attractive to use a Bayesian alternative. This is done by choosing a prior distribution for the parameter θ that has to be estimated; subsequently the normalized product of the likelihood / available information and the prior distribution is called the posterior distribution of θ. This posterior distribution tells us about the behavior of θ given the observed information and our prior beliefs about θ. The idea is that with each update this posterior distribution will remain in the same family of distributions. Azoury`s results hold for common distributions like the gamma, uniform, Weibull and the normal distribution. A point of attention is that negative effects can occur when the initial distribution function is ill-chosen, e.g. it will take longer before convergence is achieved.

In [3], where Azoury compared optimal ordering levels of Bayesian and non-Bayesian inventory models on reparable inventory systems, the most important result was that the order quantity of the Bayesian policy was always less than or equal to the non-Bayesian one. According to Azoury these lower order quantities demonstrate that in general Bayesian policies are more flexible.
Furthermore she discusses an important limitation of Bayesian models. Namely, a key assumption that the Bayesian models make is that the underlying probability distributions, although unknown, are not changing over time. Intuitively one can sense that many real-life demand processes will have unknown and changing distributions.

When we return to our role of a shop manager we want to know how we can apply this Bayesian approach to our simple model. Suppose we choose the Beta-distribution as our initial / prior distribution, which is an optimal decision according to De Groot. After observing demand k we can now update our new Beta-distribution with parameters $(r+k, s+1)$.
In our modeling step this will look as following:

$$V(x,r,s) = \max_{a \in [0,\ldots,m-x]} \left\{ \int_0^1 f(r,s)(\theta) \sum_{k=0}^{\infty} \theta(1-\theta)^{k-1} [\min(x+a,k)w - b(a) - (x+a)h + \alpha V(x+a-k, r+k, s+1)^+] d\theta \right\}$$

$with:$

$$f(r,s)(\theta) = \frac{(r+s+1)!}{r!s!} \theta^r (1-\theta)^s .$$

3.2.2 Stochastic Approximation

A second common approach to deal with unknown demand distributions is Stochastic Approximation (SA). The article I used to become familiar with this technique is "Analysis of adaptive step-size SA algorithms for parameter tracking" by Kushner and Yang [14].

They pose the problem of tracking the best parameters for a regression fit of a non-stationary process whose statistics are varying as one of the classic problems in control theory. In such cases SA algorithms can be used. The basic concept of an SA algorithm; given some unknown parameter θ, the general form of an SA procedure will be:

$$\theta_{t+1} = \theta_t + \varepsilon H(\theta_t, d_t) + \varepsilon^2 \Gamma(\theta_t, d_t) \quad with\ \forall \varepsilon > 0 .$$

In words, the new value θ_{t+1} is determined by the three parts of the right hand side of this formula. First of all it depends on the current value of θ, namely θ_t. The product in the second part determines in which direction the current θ is adjusted. Two components are used here, the step size ε and derivative function $H(\theta_t, d_t)$. $H(\theta_t, d_t)$ is the derivative function of (in our case) the currently observed demand d_t and θ_t that makes sure that θ_{t+1} is moved in the right direction. We exemplify this with the graph of a very simple time series in figure 3.1.

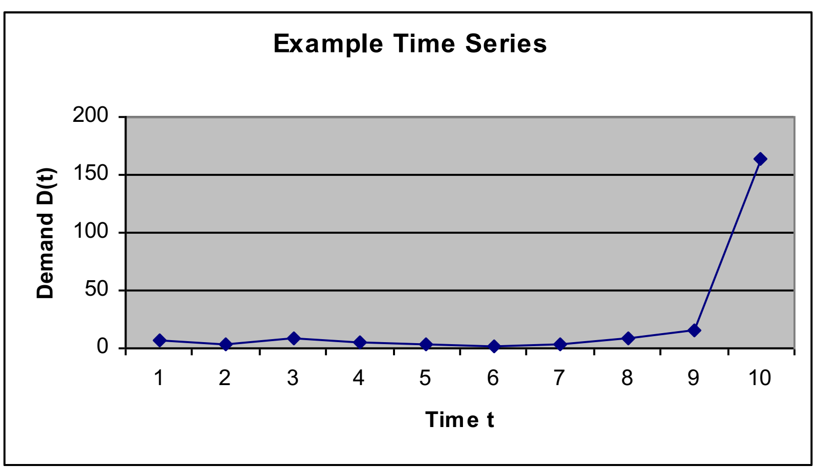

Figure 3.1 Example time series

In this example we see that from time = 8 the demand increases very fast, which means that the derivative will also be high. The idea is that stochastic approximation uses this derivative that it can rapidly come to a reasonable estimate.
The last part uses a Γ function of d_t and θ_t to generate some randomness in the estimates for global convergence, which is mostly used in engineering applications. The factor ε^2 makes sure that the influence of this Γ-function will be small enough.

An important question is what the appropriate value for ε should be. The optimal value depends on several issues such as the details of the model and the variation of its statistics $\{\theta_n\}$. In [14] proofs and data are presented for adaptive step-size algorithms for tracking these time-varying parameters when recursive stochastic approximation algorithms are used. The idea is to adjust the step size with an adaptive scheme, by replacing a fixed ε with ε_t and letting ε_t converge to an optimal value. When dealing with time-varying it is important not to let ε_t converge to 0.

One of the best known examples of an SA algorithm is the Newton-Raphson method:

$$x_{t+1} = x_t - \frac{f(x_t)}{f_t^{'}(x_t)} \quad (f_t^{'}(x_t) \neq 0)$$

In this case, problems can arise when we are confronted with a demand distribution on which information of the derivative function is unknown. In such cases we have to find our resort in methods for derivative estimation, such as Infinitesimal Perturbation Analysis (IPA). We shall not elaborate on this subject, but for the interested reader we refer to [9], in which an overview of derivative estimation techniques is given.

Nevertheless SA algorithms have the advantage of offering great freedom in their use in practical applications, e.g. suppose we want to make an SA algorithm to calculate and update the estimate for an average ψ_n that consists of i.i.d. observations ξ_i. First we define $\psi_n = \sum_{i=1}^{n} \xi_i$. Our next estimate ψ_{n+1} will be $\frac{n\psi_n + \xi_{n+1}}{n+1}$. Because we assume i.i.d. this is equivalent to $\frac{(n+1)\psi_n + \xi_{n+1} - \psi_n}{n+1}$, in which we can now recognize the familiar form of our SA algorithm $\psi_{n+1} = \psi_n + \varepsilon_n(\xi_{n+1} - \psi_n), \quad with: \varepsilon_n = \frac{1}{n+1}$.

Another familiar and widely used example that has properties of an SA algorithm is simple exponential smoothing, which is defined as: $S_{t+1} = \alpha \cdot D_t + (1-\alpha)S_t$. In this method D_t is the currently observed demand and S_t the previous demand estimate. The smoothing factor α determines in which way both components influence the forecast. One can imagine that many other update problems can be approached in a similar way.

Now let us return to our illustrative problem and see how we can apply stochastic approximation to our inventory model. We assume that our system is time-varying; therefore we shall not pay attention to the third part of the general SA procedure, because we are not expecting global convergence. This results in the following equations:

$$V(x,\theta_t) = \max_{a\in[0,\ldots,m-x]} \{\sum_{k=0}^{\infty}(1-\theta_t)^{k-1}\theta_t[\min(x+a,k)w - b(a) - (x+a)h + \alpha V(x+a-k,\theta_{t+1})^+]\}$$

$$with:$$

$$\theta_{t+1} = \theta_t + \varepsilon_t \mathrm{H}(\theta_t, \mathrm{k})$$

At the end of this section we mention a typical difference between an SA approach and a Bayesian approach is that in the Bayesian case a distribution is estimated, which ensures that all information is maintained in the estimate whereas an SA algorithm gives a point estimate and can deal with time varying parameters. The consequence is that a Bayesian method is more likely to be disrupted by its past.

3.2.3 Recent advances in Adaptive Inventory Control

One of the most interesting and few recent publications I found on Adaptive Inventory Control were written by Sox and Treharne [19, 21] in 2000 and 2002. The first article contains a literature study that specifically examines the developments in the field of Adaptive Inventory Control. In this article inventory control problems are categorized in four classes:

- fully observed, stationary demand
- fully observed, non-stationary demand
- partially observed, stationary demand
- partially observed, non-stationary demand

The authors state that a lot of information is available on the first two classes. On the class of partially observed and non-stationary demand problems previous research achievements are scarcer. The focus of their articles, as well as this paper, is mainly on the fourth category. Since many of my references coincide with the articles reviewed in their literature study, we have an indication that the techniques we discussed in the previous sections are still common approaches for this category.

Apart from the literature study the publications contain results of their own research. The key idea of Sox and Treharne`s work is to distinguish different probability distributions for different behavior in demand and associate these with different states of a Markov Decision Process. The techniques that are used to estimate the demand distributions were discussed in the previous sections. Since the focus of this paper is rather limited we shall not enter into details on Markov theory. For readers that are interested in inventory models based on Markov theory we refer to [11, 13, 17, 19, 21].

3.3 Artificial Intelligence methods

In the last decades Artificial Intelligence methods have become widely accepted problem solving techniques. Thus it will not come as a surprise that these techniques are also used in Inventory Management applications. When it comes to IM the most promising techniques are Neural Networks and Reinforcement Learning.

For this paper we have made the choice to focus on Neural Networks without elaborating on Reinforcement Learning. For readers that are interested in information on Reinforcement Learning and its applications in IM we refer to [6, 16, 18, 20], where Mitchell [16] is an excellent source for newcomers in this field. In the next sections we discuss the Neural Network in general and the specific types we chose to use for our experiments.

3.3.1 Basics of Neural Networks

One of the best known techniques in Artificial Intelligence are (artificial) Neural Networks. A Neural Network (NN) can be seen as a system that improves itself while it learns from new examples. Many modern applications use NN`s for a variety of purposes. Among these is the forecasting of demand processes. Since NN`s are applied on such a wide scale examples can be found nearly anywhere, see references [4, 7, 8, 11, 15, 16, 18, 22].

The study of artificial neural networks has been inspired by the analogy with biological learning systems that are built of highly complex networks of interconnected neurons. A single neuron uses a simple activation function to transform a vector of inputs with a given weight vector into an output neuron. In the network, the neurons are organized in a multi-layered structure, which usually consists of an input, hidden and output layer. Each layer consists of individual neurons. The specified interactions between the layers give the NN its fundamental characteristics and behavior. The task is to train a weight vector in such a way that the network minimizes the target error function whose task it is to generate an output which is as close as possible to its actual value.

According to Mitchell [14], problems that can be handled by a common Neural Network implementation have the following characteristics and advantages:

- Instances are represented by many attribute-value pairs.
- The target function output may be discrete-valued, real-valued, or a vector of several real- or discrete–valued attributes.
- The training examples may contain errors, which makes NN`s robust to noise.
- Long training times are acceptable. These times can depend on different factors such as the number of weights, the number of examples and settings for the various learning parameters and computational power.
- Fast evaluation of the learned target may be required.
- The ability for humans to understand the learned target function is not important, which makes it unnecessary to have great system knowledge.

In figure 3.2 we see an example of an NN architecture, where the NN is used to predict sales in the steel business (note that given such a sales estimate it is easy to translate it into a corresponding demand estimate.) Figure 3.2 shows us one of the great strengths of a NN, which is that we are able to choose a variety of factors from the past that affect demand. These factors serve as an input and are all used to contribute to the final demand estimate.

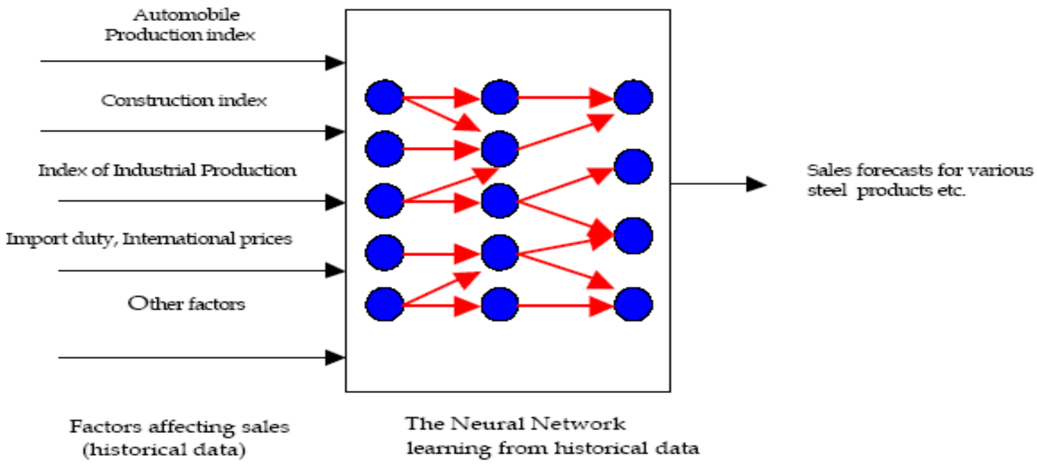

Figure 3.2 Neural Network model for prediction of steel demand [23]

We have just seen an example where a NN is used as a time series forecaster. In this case it concerned making a market prediction; other examples can be found in financial, meteorological, network traffic forecasting etc. A common NN modeling approach that is used in such cases is time windowing.

3.3.2 Time series and Feed Forward Neural Networks

Before continuing with the specific technique of time windowing we define our working environment a bit more formal. My guide for the next two sections was the article by Dorffner [8]. We shall regard our demand process as a time series, so we start by defining demand as a sequence of vectors depending on (discrete) time t : $\vec{d}(t),\ t = 0,1,\ldots$. In our case demand is only one component so we are dealing with a univariate time series. What we want is to predict future values of $\vec{d}$, which can be described formally as follows: Find a function $F:\ \Re^{n+k} \to \Re$ such as to obtain an estimate $\hat{\vec{d}}(t+p)$ of $\vec{d}$ at time $t+p$, given the values of $\vec{d}$ up to time t (plus some additional time-independent variables (exogenous features) π_i):

$$\hat{\vec{d}}(t+p) = F(\vec{d}(t),\ldots,\vec{d}(t-n+1),\pi_1,\ldots,\pi_k) + \vec{\varepsilon}(t). \qquad (3.3.1)$$

In the definition above, p is called the lag for prediction. In most cases $p = 1$, this means that the subsequent vector is estimated. This might just as well have been a point further in the future.

Furthermore we included a residual error $\vec{\varepsilon}(t)$, which is practically always adopted in a time series model to correct for measuring errors and other unknown or uncontrollable influences. For almost any real-world application it is not realistic to assume that an exact model can be found. Usually the error is assumed to be the result of a noise process, i.e. produced randomly by an unknown source. Although the noise $\vec{\varepsilon}(t)$ cannot be included into the model explicitly, many methods assume a certain characteristic of the noise (e.g. Gaussian white noise). In this way the main describing parameters (e.g. mean and variance) can be included in the modeling process. By doing this in fore-casting, one is not only able to give an estimate of the forecast value, but also an estimate of the impact of the noise on this value. This is the focus of another technique, the so-called ARCH models. For simplicity we shall neglect the additional variables π_i and $\vec{\varepsilon}(t)$ in the remainder of this paper. However it is important to keep in mind that the inclusion can be decisive in particular applications.

When we look at our demand process from this perspective, forecasting becomes a problem of function approximation, where the chosen method is to approximate the function F as close as possible. In this way, it can be compared with function approximation or regression problems involving static data vectors and many methods from that domain can be applied here. Among these methods are Feed Forward Neural Networks (FFNN), mostly applied for classification / pattern recognition and function approximation / regression. In the latter case, the FFNN can be applied to learn non-linear functions (regression) and in particular functions whose inputs form a sequence of measurements over time (time series).

The most familiar of this class of Neural Networks are Multilayer Perceptrons (MLP) and Radial Basis Function networks (RBF). Suppose we want to approximate the function $F(\vec{d}): \Re^n \rightarrow \Re^m$, a formal definition of the MLP is:

$$F^{MLP}(\vec{d}) = \left(\sum_{j=1}^{q} w_{jl}\sigma\left[\sum_{i=1}^{n} w_{ij} d_i - \theta_j \right] - \theta_l \right) \quad , l = 1,...,m \qquad (3.3.2)$$

where σ is the sigmoid function (or any other appropriate function), q the number of hidden units, w_{jl} and w_{ij} weights, and θ_i thresholds.
The RBF is defined similarly:

$$F^{RBF}(\vec{d}) = \left(\sum_{j=1}^{q} w_{jl}\Gamma(\sum_{i=1}^{n} \left\| w_{ij} - d_i \right\|) - \theta_l \right) \quad , l = 1,...,m . \qquad (3.3.3)$$

where the difference with the MLP is that the hidden units q use a distance measure and Γ is a Gaussian (or any other appropriate) transfer function. Provided that the number of units q is chosen large enough, both types have been proven to be universal function approximators. This means they should be capable to approximate any reasonable function F arbitrarily close. The approximation of non-linearity is taken into account by a superposition of several instances of the basis function.

3.3.3 Time windowing

A classical approach to deal with time series is modeling them as linear autoregressive (AR) models. This type of modeling assumes that the function F from equation (3.3.1) can be seen as a linear combination of p previous values of $\vec{d}$, which can be written as:

$$\vec{\hat{d}}(t) = \sum_{i=1}^{p} \alpha_i d(t-i) + \varepsilon(t)$$

or as a function F^{LAR} of p previous values:

$$\vec{\hat{d}}(t) = F^{LAR}(d(t-1),...,d(t-p)) + \varepsilon(t).$$

The number p determines the order of the AR model. An AR[p] model is obtained by choosing an appropriate p and an estimation of the coefficients α_i, found with some optimization method e.g. a least squares, Yule-Walker or the Burg method.

Linear autoregressive (AR[p]) modeling is naturally limited since it assumes stationarity and that the relationships between sequential elements are linear. This means we would have a problem for our non-stationary demand problem when we are tied up to just this method. Fortunately this is not the case as we can overcome these limitations by replacing the linear function F^{LAR} with an arbitrary non-linear function F^{NN} (where NN can be an MLP or RBF implementation of the NN):

$$\vec{\hat{d}}(t) = F^{NN}(d(t-1),...,d(t-p)) + \varepsilon(t).$$

The input method for this network is known as time-windowing, since it only looks at a limited part of the time series. This is clearly visible in the NN-architecture of Figure 3.3.

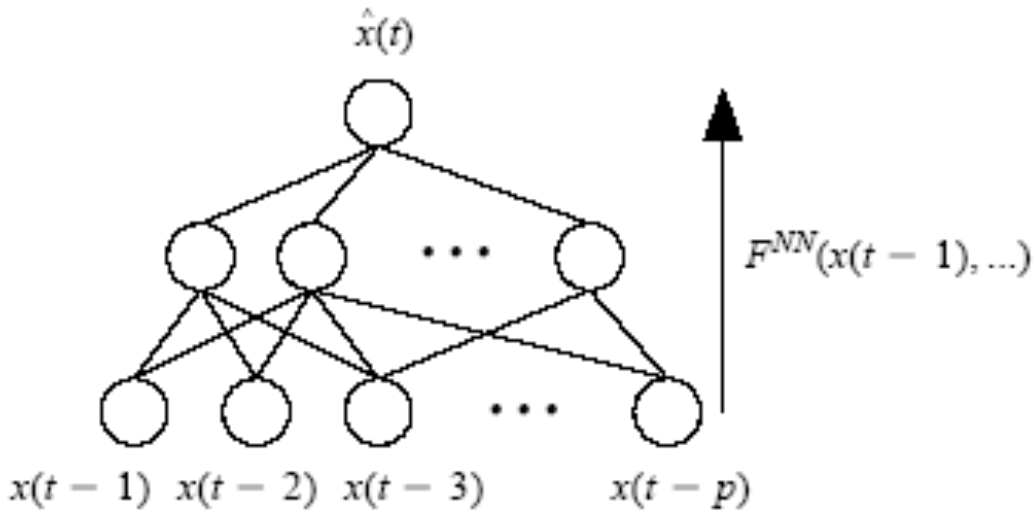

Figure 3.3 Feed Forward Neural Network model with time window [8].

Advantages of non-linear autoregressive models over linear models are that they:

- can model much more complex underlying characteristics of the series
- theoretically do not have to assume stationarity

However, we have to retain some care and cautiousness with the application of non-linear methods. Compared with linear methods disadvantages can be that they:

- require large numbers of sample data, due to their large number of degrees-of-freedom
- can run into a variety of problems, such as overfitting, sub-optimal minima as a result of estimation (learning), etc.
- do not necessarily extend the linear case in a trivial way

The first point is especially important for many real-world applications, where it is often the case that only limited amounts of data are available. Therefore in many cases a linear model might still get the preference, even when it is known that dependencies are non-linear. The second point concerns the type of learning algorithm. These problems can be more severe than in the linear case, where a phenomenon as overfitting can be avoided by for instance choosing the value for the parameter p too high.

4. Experiments

In the previous chapter we have seen different techniques to forecast future values of demand when the demand process follows an erratic course. In this chapter we shall try to develop a feeling for the accuracy of these techniques.

By analogy with the article of Carmo and Rodrigues "Adaptive Forecasting of Irregular Demand Processes" [4], where comparisons are made between relatively simple exponential smoothing methods and radial basis function neural networks, we too made a (simple) experimental set-up. The question we are asking ourselves is straightforward;

Which method is most suited to make one-step-ahead predictions when it is confronted with a univariate time series of an arbitrary real-life process?

In this experiment we shall not so much look at the shape of the data, but focus on determining the generally most accurate forecasting method of a group of very simple techniques. These techniques are described in paragraph two. Since it would be nice to make some sort of generalization for our results we tested the selected techniques on three different data sets of different origins. The datasets are described in the first paragraph. In paragraph three the results are presented and discussed.

4.1 Data

The datasets for this experiment are obtained from the following three different real life sources:

4.1.1 Dataset 1: Printer volumes

The first dataset concerns the printer volumes of a certain individual. For each job individual volumes were summarized on a daily level. These totals result in the first time series of size: n = 216.

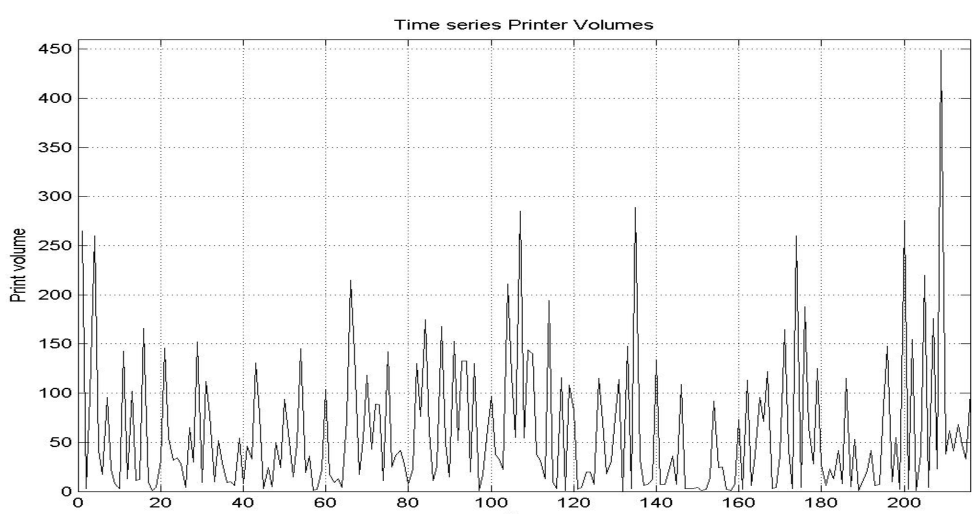

Figure 4.1 Time series printer volumes.

4.1.2 Dataset 2: Hits Computer Science VU

The second dataset is obtained as following: we collected the log files with timestamps for approximately 20 days (these are logged in a file on the Apache-server) of each time the website http://www.cs.vu.nl/ was visited. Because this concerns ≈ 4.8 million individual records, a script was written to make aggregations so that the dataset became more tractable. In this case the minute was chosen as the level of aggregation, which resulted in a time series of n = 27768.

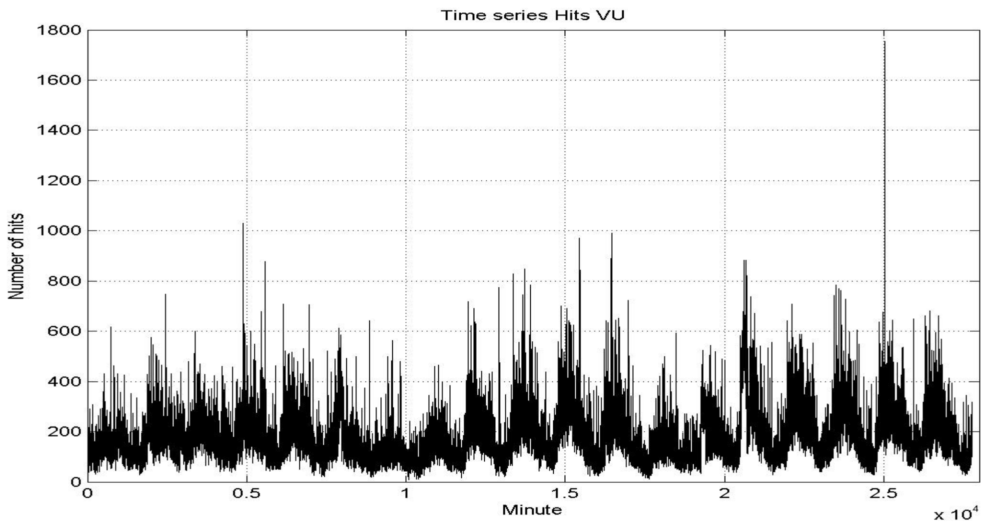

Figure 4.2 Time series hits www.cs.vu.nl.

4.1.3 Dataset 3: Send times data packages

The last dataset that is used was provided by Menno Dobber. The data concerns send / receive times from small data packages between a server from the VU and a server in Taiwan (n = 20000).

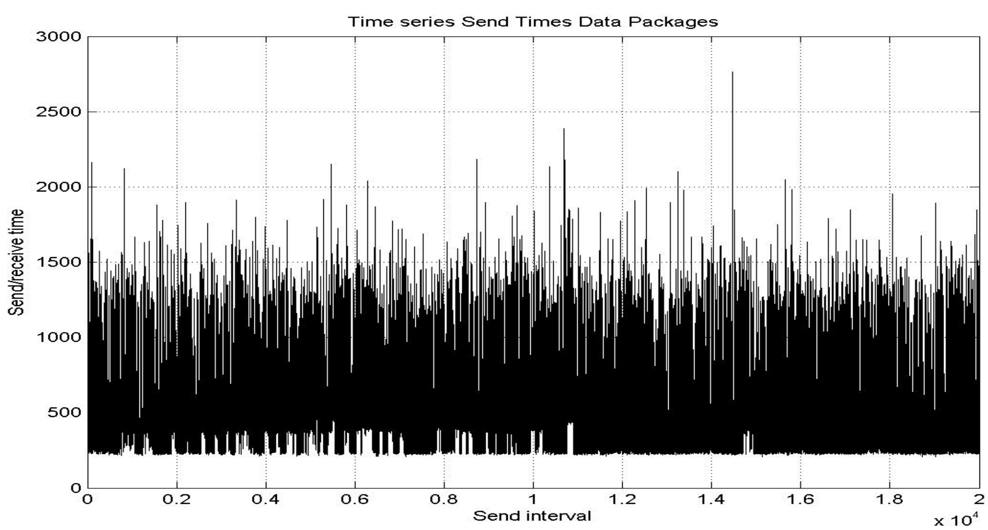

Figure 4.3 Time series Send times data packages.

4.2 Methods

The methods we shall use are very basic since the main idea behind these experiments is to develop a feeling for the performance of the different forecasting methods. We made a selection of 5 methods (some with adjustable parameters) to estimate the value of demand D at time $t+1$, these are:

1.) Moving average (n) (not to be confused with MA-processes):

$$\hat{D}_{t+1} = \frac{1}{n} \sum_{i=t-n+1}^{t} D_i \quad \textit{with } n = 1,\ 2,\ 5,\ 10. \tag{4.2.1}$$

2.) Simple exponential smoothing:

$$\hat{D}_{t+1} = \alpha_i \cdot D_t + (1-\alpha_i)\hat{D}_t, \quad \textit{where } \alpha_i = \tfrac{i}{20} \quad \textit{and } i = 1,\ldots,20 \tag{4.2.2}$$

3.) Updating mean:

$$\hat{D}_{t+1} = \frac{(t-1)S_t + D_t}{t} \tag{4.2.3}$$

Notice that the nature of this method makes: $S_t = \sum_{i=1}^{t-1} d_i$.

4.) MLP Neural Network (with time window *n*)

$$\hat{d}_{t+1} = \left[\sum_{j=1}^{q} w_j \sigma\left(\sum_{i=t-n+1}^{t} w_{ij} d_i - \theta_j \right) - \theta \right] \tag{4.2.4}$$

5.) RBF Neural Network (with time window *n*)

$$\hat{d}_{t+1} = \left(\sum_{j=1}^{q} w_j \Gamma\left(\sum_{i=t-n+1}^{t} \left\| w_{ij} - d_i \right\| \right) - \theta \right) \tag{4.2.5}$$

Obviously the last two methods are far more sophisticated than the other three (note that for these two the necessity for scaling and rescaling is made insightful by using the notation d instead of D). But since any comparison has to start somewhere and for the time being it is unknown whether these fancy methods will actually perform better than the very simple methods. For example: Method (4.2.1) will do fine when $\hat{d}_{t+1}$ can be estimated as an average over the previous n values (a limited time window), (4.2.3) when $\hat{d}_{t+1}$ does not differ much from its mean, whereas variation of α in method (4.2.2) allows a balance between the previous value and the smoothing component also looks as a method that might perform quite well.

Method (4.2.1) and (4.2.2) can be found in a book on the basics principles of logistics by Visser and van Goor [22]. Method (4.2.3) is a test of our very simple SA procedure from Chapter 3.2.2. The last two methods were the Neureal networks discussed in Chapter 3.3.1.

This brings us to the issue of performance measurement. We chose to use the most common method, the root mean square error: $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}[D(t-i) - \hat{D}(t-i)]^2}$.

Other setup issues concern questions like; should the data be pre-processed, and should the data be scaled or not. We used no pre-processing because our assumption in this project is that an algorithm should work fast and accurate on any real-life dataset without a lot of preparatory work. In practice, and when there is more time available, one might prefer to spend more time to analyze the process at hand. To use the NN properly we scaled our original vector $\vec{D}$ onto $\vec{d} \in [-0.8,\ldots,0.8]$, according to the familiar scaling rule:

$$d_i = 0.8\left[2\left(\frac{D_i - \min(D_i)}{\max(D_i) - \min(D_i)}\right) - 1\right], \quad i = 1,\ldots,n.$$

After applying the network we rescaled to the original scale.

From some methods we can tune parameters by varying:

- (4.2.1) the window size.
- (4.2.3) the vales of α .
- (4.2.4) / (4.2.5) the window size, number of hidden units, activation functions and optimization method.

The results in section three shall contain a mixture of the most noticeable combinations we found by tuning the parameters of each method.

4.3 Results

Before presenting the results we take some time to become more familiar with our datasets (see Figure 4.4). What catches the eye most is that all datasets are distributed to the left and have fairly high variances. These properties make it extra interesting to see whether the performance of the forecasting methods is acceptable.

Dataset/ RMSE	Hits VU	Printer volumes	Send times
N	27768	216	20000
Mean	175	60	474
Standard deviation	100	69	336
Minimum	8	0	205
First quantile	104	9	246
Median	153	33	334
Third quantile	221	95	539
Maximum	1754	449	6800

Figure 4.4 Standard summary of each dataset.

4.2.1 Validation

To validate the results we compared the mean of each forecasted vector with the original mean in Figure 4.4. Whenever this differs a lot it is an indication that the forecasting method is not constructed properly.

Another validation method was to compare the mean square errors of each method with each other. As much as we would like to find a perfect forecasting method, we have to assume that the results will not differ a great deal from each other. To reduce the chance of any negative effects during the start-up phase, we gave all methods a warm-up period by ignoring the first m forecasts.

Finally, we chose not to use any cross-validation method (dividing the data in test and train sets). Therefore the results of the Neural Networks shall contain so-called resubstitution errors. This is unavoidable because if we would use a cross-validation strategy we would not make a fair comparison since all other methods are making their forecasts with more recent observations.

4.2.2 Results of the experiments

Dataset/ RMSE	Hits VU	Printer volumes	Send times
RBF (Gaussian)	67	63	329
RBF (tps)	71	73	329
RBF (r4logr)	71	74	329
MLP (linear)	68	64	327
Last	81	118	436
Moving average 2	75	90	376
Moving average 5	72	79	356
Moving average 10	71	79	344
Updated mean	100	77	333
BestExp	69	78	330
Best α	0,20	0,05	0,05

Figure 4.5 Results / MSE`s of each different method for all datasets.

Dataset/ Rank	Hits VU	Printer volumes	Send times	Sum of ranks	Total rank
RBF (gaussian)	1	1	2	4	1
RBF (tps)	4	3	3	10	3
RBF (r4logr)	5	4	4	13	4
MLP (linear)	2	2	1	5	2
Last	9	10	10	29	10
Moving average 2	8	9	9	26	9
Moving average 5	7	8	8	23	8
Moving average 10	6	7	7	20	6
Updated mean	10	5	6	21	7
BestExp	3	6	5	14	5

Figure 4.6 Ranks of each different method for the three datasets.

In Figure 4.5 we find the results of our simple experiment. It is clear that the most sophisticated techniques, the radial basis function and MLP neural network (with different activation functions), have produced the best results. The other results are from: the different moving averages, the updated mean and exponential smoothing (only the results with the coefficient that performed best).

We gain a clear insight into these mutual performances by looking at Figure 4.6, where we see a table containing the individual ranks for each dataset, the sum of these ranks and the overall rank.

One of the things that is worth mentioning is that when we compare the RMSE`s for each method - dataset pair with the standard deviation of each dataset, the relative difference between these values are quite small in a large number of cases.
This means that the percentage of explained variance (or fit) will also be quite low and consequently the involved model is inaccurate. In our case an estimate of the fit can be obtained by: $\hat{R}^2 = 1 - \frac{RMSE^2}{\sigma^2}$.

Since the idea of this experiment is to gain a feeling for our methods in a very sober setup, we suffice with noticing these modest fits and point out that these results only served an indicative purpose. We leave it for further studies to analyze better setups and more sophisticated methods that deliver better fits. And even then it is quite conceivable that an overall conclusion should be that the data contains little information. Not all real life data deserves the attention of a statistician.

5. Discussion

In this discussion we confine ourselves to the two main focuses of this paper.

1. The future status and use of intelligent forecasting applications in Inventory Management.

Even though the leaders in business intelligence are developing interesting and valuable applications, our notion is that they seem to have difficulties to let them come off the ground. In several articles on business intelligence we can read between the lines that only a very small percentage of companies and/or organizations have actually implemented such applications. This feeling is confirmed by insiders such as account managers of companies like SAS. The cause for this lack of use has various explanations. The first is that these applications are often relatively expensive for a company when little is known in advance about the financial gain of the investment (especially since we find ourselves in a time of world-wide economic malaises). Other explanations can either be ascribed to the fact that internal information management systems are not (yet) equipped to support such sophisticated applications or simply because the people in the organization are not familiar or do not see the surplus value for their business. The challenge for the future is to open the eyes of these organizations by pointing them on the attractive financial prospects. It is my opinion that also papers like this one can fulfill a role in contributing to that familiarity and awareness.

2. Development and comparison of adaptive forecasting methods in the future.

One of my first findings in this study was that it appeared to be quite difficult to obtain recent literature on adaptive forecasting techniques in Inventory Management. In my search I was looking for a survey that could give insight in the development of forecasting techniques, but also presented an overview of (recent) techniques that could easily be put in practice. However, most of the articles I encountered could not be divided into one of these categories. Either they were purely focused on giving an abstract mathematical proof of a solution for a certain problem, or it were stories to describe a successfully implemented technique as a solution to a very specific practical problem. Of course it is not strange that little can be found in such a small part of the total area of forecasting methods, and that it seems that most of the encountered techniques are available for quite a while.

This made us wonder in general whether today's forecasting methods are actually better than in the past. We found answers to this question in a study by Franses [8], performed in 2003. He carried out a survey amongst academics working in the field of forecasting. The main objective of his study was to find out if people working in this discipline felt that significant progress has been made in the last three decades. A conclusion was that only modest progress was observed, although the profession seems far from satisfied. This progress appears to be mainly due to the increase in computational power and the fact that we are better capable to incorporate important data features in our models. According to this article progress could have been faster if we would have been able to include expert opinions. The last two form important topics on the research agenda.

Another direction from which future advances could be expected is the work done on the intersection of conventional mathematical and artificial intelligence methods. In the currently available literature it remains unclear which of these approaches deserves recommendation. With the article by Carmo [5], we found just one recent article that made actual comparisons between (basic) mathematical and artificial intelligence techniques. An explanation for this lack of comparisons is that, traditionally these two schools approached problems from different perspectives. During this investigation we learned that the typical contrasting aspects between the schools are diminishing, so it is likely and desirably that more comparisons will follow in the near future.

The lack of comparisons made us feel challenged to setup our own very basic experiments with interesting real life datasets. The primary focus was to develop a feeling for the performance of our selection of methods. In these experiments the neural networks were the obvious winner. Although the results are of modest importance from a scientific point of view, it provides a basis and hopefully encourages people to do further work in comparing mathematical approaches with modern learning techniques. A future step is to make comparisons between methods with greater predictive qualities.

6. References

6.1 Articles

1. C. Abraham, V. Ahuja, A.K. Ghosh, P. Pakanati. "Inventory Management using Passive RFID Tags: A survey". University of Texas, Department of Computer Science, 2001.

2. K.S. Azoury. "Bayes solution to dynamic inventory models under unknown demand distribution". Management Science, 31:1150--1160, 1985.

3. K.S. Azoury, B.L. Miller. "A comparison of the optimal ordering levels of Bayesian and non-Bayesian inventory models". Management Science, 30:993--1003, 1984.

4. J.L. Carmo, A.J. Rodrigues. "Adaptive Forecasting of Irregular Demand Processes", Engineering Applications of Artificial Intelligence, 17, 137—143, 2004.

5. Y.F. Chen, J.K. Ryan, D. Simchi-Levi, "The Impact of Exponential Smoothing Forecasts on the Bullwhip Effect", Naval Research Logistics, No. 3, 2000.

6. R.H. Crites, A.G. Barto, "Improving Elevator Performance Using Reinforcement Learning". Advances in Neural Information Processing Systems 8, MIT Press, Cambridge MA, 1996.

7. F.R.J., Davey N., Hunt S.P., "Time Series Prediction and Neural Networks". Journal of Intelligent and Robotic Systems 31:91-103, 2001.

8. G. Dorffner. "Neural networks for time series processing". Neural Network World, 6(4) 447-468, 1996.

9. P. L' Ecuyer. "An overview of derivative estimation". Proceedings of the 1991 Winter Simulation Conference, 207 – 217, 1991.

10. P.H. Franses. "Do we make better forecasts today? A survey amongst academics". Econometric Institute Report EI 2003-06, Rotterdam 2003.

11. P.H. Franses. Time series models for business and economic forecasting. Cambridge University Press (UK), 1998.

12. S. Karlin. "Dynamic inventory policy with varying stochastic demands". Management Science, 6:231--258, 1960.

13. P.R. Kumar. "A survey of some results in stochastic adaptive control". SIAM Journal of Control and Optimization. 23:329--380, 1985.

14. H.J. Kushner, J. Yang. "Analysis of Adaptive Step-Size SA Algorithms for Parameter Tracking". IEEE transaction on Automatic Control, vol. 40, N.8, 1403-1410, August 1995.

15. H. Lee. "Simple theories for complex logistics". Optimize, Issue 22, July 2004.

16. T.M. Mitchell. Machine Learning. McGraw-Hill 1997.

17. T. Odanaka. "Analytical and computational solution of adaptive inventory processes". Journal of Mathematical Analysis and Application, 105, 405--418, 1985.

18. B. van Roy, D.P. Bertsekas, Y. Lee, J.N. Tsitsiklis, "A Neuro-Dynamic Programming Approach to Retailer Inventory Management". Lab. for Information and Decision Systems Report, Nov. 1996.

19. C.R. Sox, J.R. Treharne. "Near-Optimal Policies for Adaptive Inventory Control with Partially Observed, Non-Stationary Demand". 2000 NSF Design and Manufacturing Research Conference, 1--16, 2000.

20. G. Tesauro, "Practical Issues in Temporal Difference Learning". Machine Learning 8, 257-277, 1992.

21. J.R. Treharne, C.R. Sox. "Adaptive Inventory Control for Nonstationary Demand and Partial Information". Management Science, 48:607--624, 2002.

22. H.M. Visser, A.R. van Goor. Werken met logistiek. Stenfert Kroese, 1999.

23. P. Yadav, Sridhar P.A.V. "Demand management in the supply chain". Proceedings of OPSCON 98, 1998.

6.2 Internet links

1. http://www.sas.com/
2. http://www.sap.com/
3. http://www.microsoft.com/
4. http://www.capgemini.com/
5. http://www.oracle.com/
6. http://www.ibm.com/us/
7. http://www.evant.com/