Harun Smailbegović, BSc

Optimisation of the Requirement Planning Process in Serial Production

Master's Thesis

to achieve the university degree of

Master of Science

Master's degree programme: Production Science and Management

submitted to

Graz University of Technology

Supervisor

Dipl.-Ing. Dietmar Neubacher

Institute for Engineering and Business informatics
Head: Univ.-Prof. Dipl.-Ing. Dr.techn. Siegfried Vössner

Graz, November 2017

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

Date	Signature

Abstract

Organizations nowadays have more data at their disposal than ever before. However, the lack of competence and readiness to use this information very often leads to the loss of profit increasing opportunities. In a world of ever-increasing demand for customizable and exclusive products, it becomes progressively difficult to predict their sales. This can further lead to an inability of accurately estimating inventory levels as well as choosing inappropriate planning and forecasting strategies. The problem becomes even more evident in the case of industries with an extensive product portfolio and high demand volatility.

The research bases itself on the case study method and investigates the potential behind using historical demand data as means of generating valuable knowledge. The study was done on the German kitchen manufacturing company ALNO AG. Series of interviews were conducted in order to familiarize with the current planning process which was later thoroughly investigated and described. Records of historical demand for 750 subassemblies from the company's serial production were collected, processed and analyzed. Based on this data, parts were classified according to the variability of their demand and their relative importance. The classification served as a guiding structure for thesis recommendations and helped with allocating planning resources to the most important parts. Computer developed algorithms were then used to employ analytical tools with the goal of examining improvement possibilities regarding demand forecasting and inventory management.

Results showed that using statistical methods is a more efficient way of predicting the future demand and can improve forecasting by 2,5%. A simulation study and regression analysis showed that the service level of 98% can be successfully reached by connecting the safety stock with the demand's coefficient of variance. This independent variable describes 61% of the variation in the required safety stock.

The main contribution of this thesis, however, is not in the series of results which the analyzed data set provided, but in explaining the actual process of looking for improvements in the area of production planning and inventory management through the application of statistics and other mathematical tools. It shows how exploiting otherwise unused data can give companies valuable insights into their processes and manufactured goods.

Contents

1 Introduction

The main purpose of running a for-profit organization is to deliver the profit for shareholders. In other words, all organizations strive to increase their revenue and reduce costs simultaneously. However, product price in today's world is largely determined by the competition, and very often the most convenient way for manufacturing companies to increase profit is to produce more efficiently.

Sule [23] states that the main goal of production planning is to increase efficiency in manufacturing and consequently create a larger profit margin for the company. However, many of the tools commonly used today were developed in times of lower competition and more stable markets. Although these tools are still very beneficial, staying competitive in today's industries requires more sophisticated approach.

Fortunately, as a result of technological progress, many companies nowadays have access to a large amount of data. What's more, data science is becoming an increasingly important tool in all areas of business including production planning and control. Many companies, however, either fail to recognize this importance or don't have the knowledge to identify improvement opportunities.

Hence, by using case study method, this thesis aims to further the understanding of employing data analytics and other mathematical tools as means of advancing the production planning process.

1.1 Motivation

MRP and ERP systems such as SAP usually don't erase the collected data. It is very often only a byproduct of some control process, and once it becomes historical, it stops adding value to the business. In such cases, companies can be unaware of its lying potential.

Gartner research [18] states in its article that 70% of captured manufacturing data remains unused today. They call this phenomenon dark data and define it as the information assets organizations collect, process and store during regular business activities, but generally fail to use for other purposes such as analytics. [17]

Analytics Advantage survey found that one of the main reasons behind the rising importance of data analytics is the fact that a substantial amount of data is still not used. [11]

Meanwhile, 33 percent of the Honeywell survey [9] participants declared that their companies have no plans to invest in data analytics in the next year. Of those 33% :

- 61 % believe their organizations already have systems in place to ensure safety, yield and success
- 45 % said their companies have seen some growth without data analytics
- 42 % said they don't fully understand the benefits of big data
- 35 % believe people are overstating the benefits of big data

Unused data usually means loss of knowledge obtaining potential. Because of the lack of attention directed towards its cleaning and maintenance after it's been stored, it is also very often of questionable quality.

Convincing stakeholders to use this data and pay more attention to it is never an easy task. Many times they will agree that data analytics is a task of strategic importance, but this rarely meets reality. For them, implementation of these concepts means an allocation of available resources to the project they are not sure will pay off. Usually, it seems more reasonable to invest time and money in something they find guaranteed to make a return.

Given these circumstances, the main motivation behind the thesis lies in the possibility of contributing to the better understanding of data analytics role in production planning and control process. A prove that large companies such as the one studied in this thesis often miss out on the improvement opportunities by not fully exploiting the available data could be an incentive for stakeholders in other companies not to make the same mistake.

1.2 Thesis statement

Records of historical demand, often left unused by companies, are an important source of knowledge necessary to plan a production in the most efficient manner. Therefore, data analytics should play a key role in the optimization of production planning and control.

1.3 Methodology

This thesis relies on the case study as the main research method. Yin [24] describes the case study as an empirical inquiry that investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used. He also emphasizes that case studies can include, and even be limited to, quantitative evidence.

It is important to mention that the research conducted in this thesis relies itself heavily on the fact that case studies are not precisely planned and very often don't have a predefined structure for the observations and analyses. This helps investigators to guide the study by they see in the field. Becker [2] explains an importance of this kind of approach:

"It prepares the investigator to deal with unexpected findings and, indeed, requires him to reorient his study in the light of such developments. It forces him to consider, however crudely, the multiple interrelations of the particular phenomena he observes. And it saves him from making assumptions that may turn out to be incorrect about matters that are relevant, though tangential, to his main concerns. This is because a case study will nearly always provide some facts to guide those assumptions, while studies with more limited data-gathering procedures are forced to assume what the observer making a case study can check on."

Conducting case studies is done for various purposes. As a research method, they are usually performed to generate findings of relevance beyond the individual case. However, it is important to mention that they are also of interest when one studies an organization with the aim of improving its functioning. [6]

This thesis is a result of three months of active participation in the company and it should be mentioned that some its findings were used as soon as they were discovered. Therefore, contribution by answering the research questions was not the only goal, in addition to this it was also important to solve the problems that have been identified within the studied company. A variant of a case study which fits this idea is a so-called action research. Same as every type of case study, action research uses evaluations of particular subjects, such as an organization, a group of people, or a system at a point of time. It attempts to capture the "reality" in greater detail and typically no control of the phenomena is exercised.

To conclude, action research differs from the typical case study in a way that it also includes participation in the area of study. The researcher simultaneously evaluates the results of this participation. In other words, the focus is placed not only on answering the research questions but also on improving the situation in the organization.

1.4 Thesis structure

Apart from the introductory part and the conclusion, this thesis consists of three more chapters which represent the research part. Here, a brief introduction to each of them will be given.

Theoretical framework

Designed in a way to provide the research base and a scientific background as a support for the methods used to conduct a study. Explanation of the theoretical framework also helps the reader understand the perspective and context of the thesis, as well as the studied case.

Case study introduction

This chapter serves as an opening to the case study. It begins with a broader view as it briefly describes the company in which the study was conducted. The focus gradually narrows with description of company's product, manufacturing process and ultimately the production planning process as the main point of interest for this research. Here the relationship between the thesis research focus and the case itself is established and defined by introducing problems which studied company faces. The chapter further describes the project in detail by defining its scope and methods used in this case study to confirm the thesis statement.

Case study research

This is the main part of the thesis where most of the data analysis is done. Different sections in it deal with different parts of the research. Every section describes the improvement opportunities in the specific domain and the analysis conducted to attain them. The following is a brief description of each of these chapters.

- **Analysis of the current requirement determination process**
  Findings explained in this section appeared as a result of the initial efforts to familiarize with the planning process. Interviews conducted with MRP Controllers concerning their job revealed a few main methods they use to make predicitons about the future requirements. Through a comprehensive analysis of these methods, some significant flaws were discovered. The intention of this section is to present these downsides and build a base for further study.

- **Data preparation**
  This section explains the process of determining the source of data used for the research. It further describes how the data was collected, processed and prepared to serve as an input for the later developed algorithms. Series of interviews and literature research also contributed to the determination of most suiting methods for data processing. This section, in and of itself, can be of great help for the company and further research projects.

- **Classification**
  Considering the number of parts included in this research, classifying them was an inevitable thing to do. This chapter describes the characteristics according to which the parts were classified, the methods of their calculation and the algorithm used for this purpose.

- **Statistical forecasting**
  This section bases itself on the flaws discovered in the presently used methods of predicting future demand and investigates possibilities of using statistical forecasting methods to improve these predictions. Accuracy, as well as the complexity and time requirements of the process, were the primary decision variables when selecting the most appropriate methods.

- **Inventory management**
  Unsuitableness of the demand data for the accurate forecasting led to the realisation of safety stock importance in the produciton planning process. Furthermore, analysis of the inventory policies revealed shortages in the safety stock determination procedure. This section, therefore, aims to prove the existance of correlation between required safety stock and properties of demand data distribution. The findings are then used to create recommendations for the company.

- **Case study conclusion**
  In order to create usefull recommendations, the key results are first briefly described. This chapter introduces findings to serve as a basis for the Conclusion chapter.

2 Theoretical framework

As the research in this thesis was to a great extent guided by what was observed in the field, it ended up covering more than one area. This includes Production Planning, Forecasting as a part of it and Inventory Management. It is important to mention, however, that these disciplines are linked very closely. Figure 2.1 shows how they relate to each other.

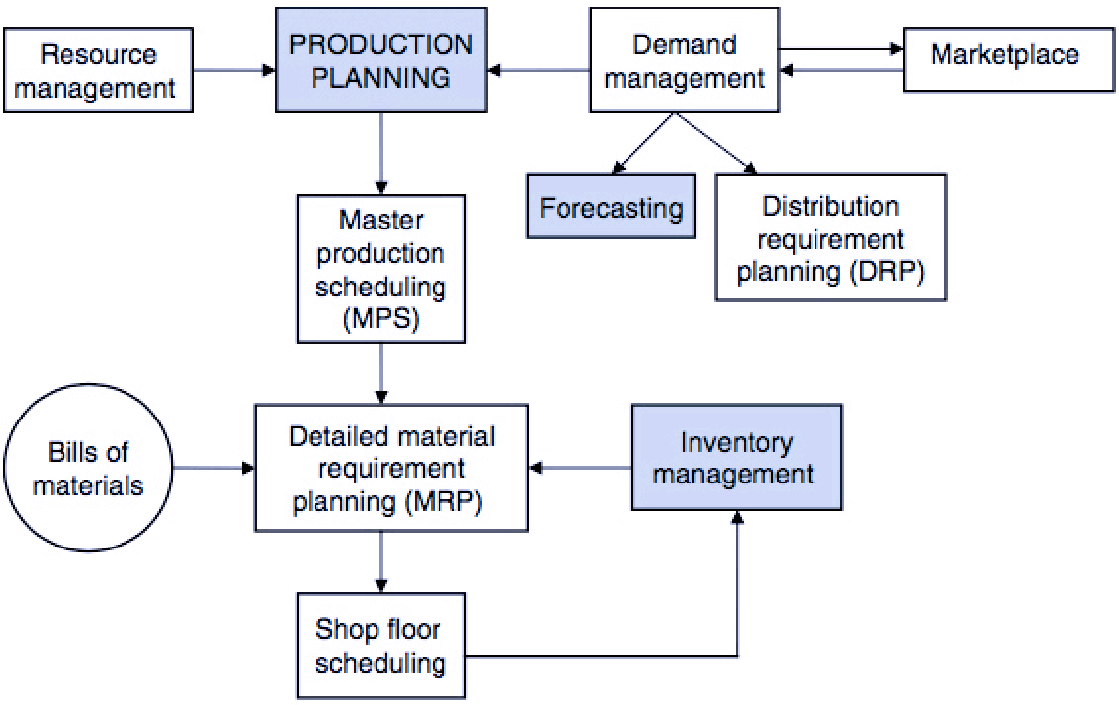

Figure 2.1: Diagram connecting topics that are discussed in the thesis.

Sections in this chapter will provide the brief theoretical introduction to these disciplines.

2.1 Production planning

The central function of organizations of all sizes is the production of some defined output from its processes. To be as efficient as possible organizations must understand and apply principles of production planning and control to this process output as it is being produced. The task of production planning is to increase efficiency in manufacturing and improve effectiveness in customer service. Production planning determines what, when, and how much to produce to meet the demand and minimize the excessive inventory or back orders. Scheduling determines how to achieve those goals when the resources are limited; and, if the goals cannot be realized, how best to find the optimum solution with the available resources. [23]

Understanding the context of this thesis requires identifying the design of the company's planning and control system. Chapman [3] states that this design depends mostly on the volume and variety of the demand, and that is generally driven by the amount of customer influence in the design of the product. The extent of it is usually described by the following categories:

- **Make-to-Stock (MTS)**
  These are products that are completely made into their final form and stored as finished goods. Customers are then supplied from an inventory and don't have much influence on the design of the product. Suppliers produce these products to their specifications and the only option customers have is to either purchase it or not.

- **Assemble-to-Order (ATO)**
  These products are assembled from components only once an order has been made. Customers have some more influence on the design, as they most often have a possibility of selecting various options from predesigned subassemblies. The key components used in the assembly are planned and usually stored in anticipation of a customer order.

- **Make-to-Order (MTO)**
  Here the customer can specify the design of the final product as long as it still consists of standard materials and components. These products are most often highly customized, produced in low volume, and their manufacturing starts only after a customer's order is received.

- **Engineer-to-Order (ETO)**
  The design of these products is almost completely in the hands of the customer. He is often not even limited to the standard materials and components. The product is created from a collaboration with the customer, beginning with a need and a concept.

2.2 Forecasting

In order for companies to decide when to produce and how much to produce, they require some type of future demand estimation. Without this, it is impossible to predict how much capacity or supply will be needed to meet this demand.

"Forecasting is a technique for using past experiences to project expectations for the future."[3]

They serve as a basis for budgeting, sales, purchasing, personnel, inventory and many other tasks. According to Chapman [3] there are five fundamental characteristics of forecasts:

- **Forecasts are almost always wrong.** It is almost never a question if a forecast is correct or not, but how wrong is it expected to be and how to accommodate the potential error.

- **Forecasts are more accurate for groups or families of items.** It is usually easier to develop a good forecast for a product line than it is for an individual product.

- **Forecasts are more accurate for shorter time periods.** Demand for extended time periods far into the future is generally less reliable.

- **Every forecast should include an estimate of error.** To be complete, a good forecast has both the forecast estimate and the estimate of the error. This error should be minimized as much as possible.

- **Forecasts are no substitute for calculated demand.** When real data is available, it should always be used over forecast.

There are two general approaches to forecasting; qualitative and quantitative, which will be introduced in the upcoming sections.

2.2.1 Qualitative forecasting

If management must create predictions quickly, they may not have enough time to collect and analyze quantitative data. Past data may also not always be available, for instance when a new product is being introduced. In these cases, forecasters very often rely only on judgment and opinion. [22]

Qualitative forecasting uses knowledge of highly experienced employees and consultants, rather than statistics. It may be less time-consuming but tends to have a bias based on the personality of the forecaster. A pessimistic person tends to overestimate, while an optimistic person has the opposite tendency.[23] The general belief is that a group of experts can make better predictions than a single person. This way the experience and knowledge of multiple individuals can be used to address the problem.

Chapman [3] mentions market surveys, Delphi consensus, life-cycle analogies, and informed judgment as some of the more common quantitative forecasting methods.

2.2.2 Quantitative forecasting

Uses historical data and analytical techniques to make predictions. It can be further divided into two classes.

- **Casual forecasting**
  Causal forecasting methods are used where there is a strong correlation between two or more variables. The essence of casual forecasting is the development of an equation which shows how independent variables influence the predicted one. Common methods used in such analyses are regression analysis, modeling and simulation, econometric modeling.

- **Time series forecasting**
  A time series is a sequence of data points, typically consisting of successive measurements or observations on a quantifiable variable(s), made over a time interval [4]. Generally time is assumed to be a discrete variable. Time series observations are usually chronological and recorded at equally spaced points in time. However, it is important to mention that intervals can also be irregular.
  Investigation of time series, generally starts with graphical representations. This allows quick detection of those properties which can be seen by simply observing the time series plot. For this purpose most often used representation is a line chart.

Time series decomposition

The simple method of describing a time series is by decomposition. According to Diggle [5], classical decomposition is of the notion that time series can be decomposed into four elements:

Trend

The trend represents a long term pattern of a time series. Depending on the direction in the data, the trend can be positive or negative. It is not always linear; sometimes it is also modeled as quadratic or exponential function.

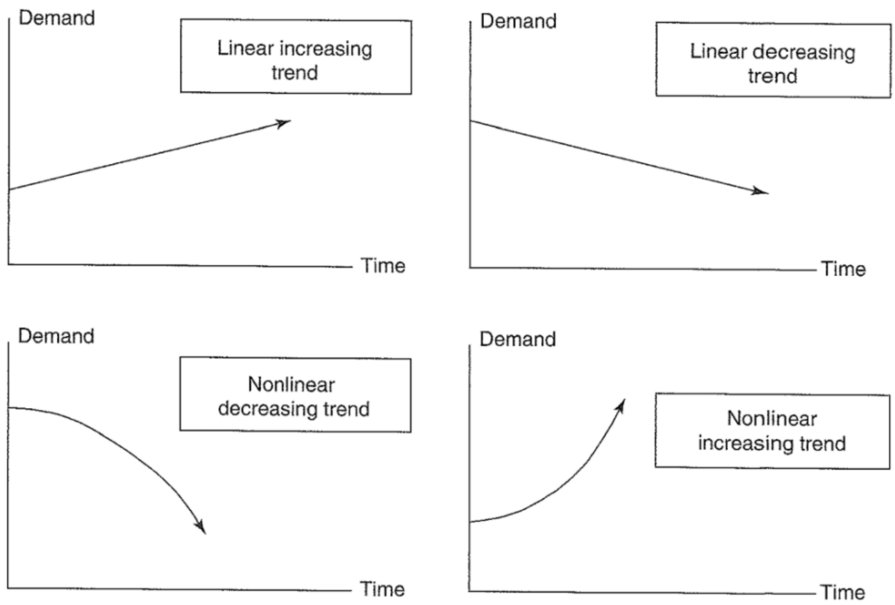

Figure 2.2: Examples of trends[3].

Seasonal effect

A seasonal pattern appears when a time series is influenced by a calendar. The period of fluctuations is always fixed. Most common types of seasonality are quarterly, monthly, weekly and daily.

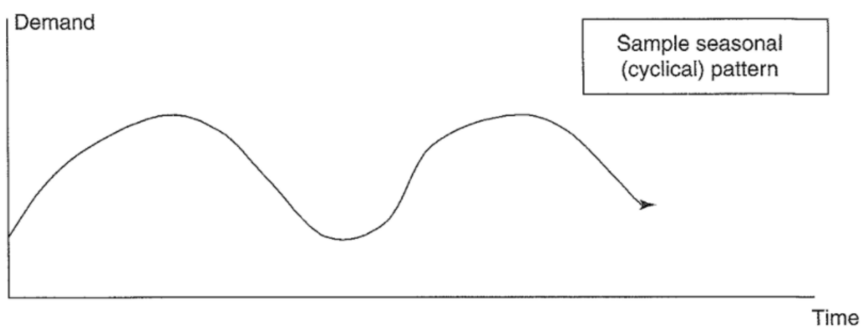

Figure 2.3: Example of seasonal pattern[3].

Cycles

This element represents cyclical fluctuations in the data. When data rises and falls around a given trend, it usually implies the presence of the cyclical pattern. The usual reason behind it are changes in the business or industry.

Residuals

Other random or systematic fluctuations. This element is unpredictable.

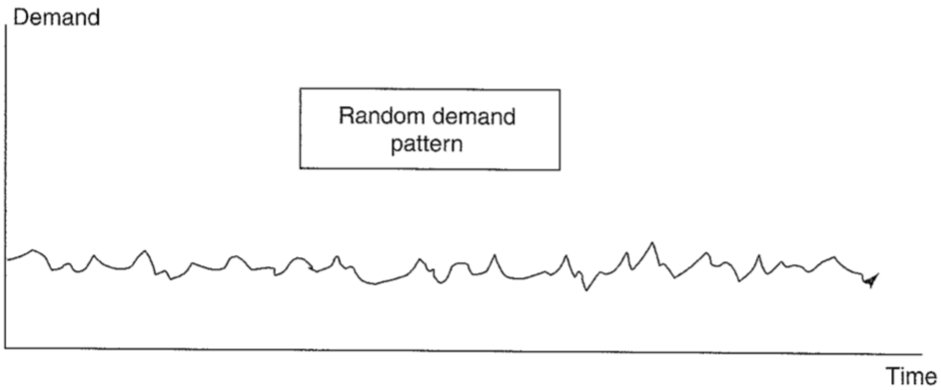

Figure 2.4: Example of random fluctuations[3].

Diggle [5] states that the idea is to create separate models for these four elements and then combine them, either additively

$$X_t = T_t + I_t + C_t + E_t,$$

or multiplicatively

$$X_t = T_t \cdot I_t \cdot C_t \cdot E_t.$$

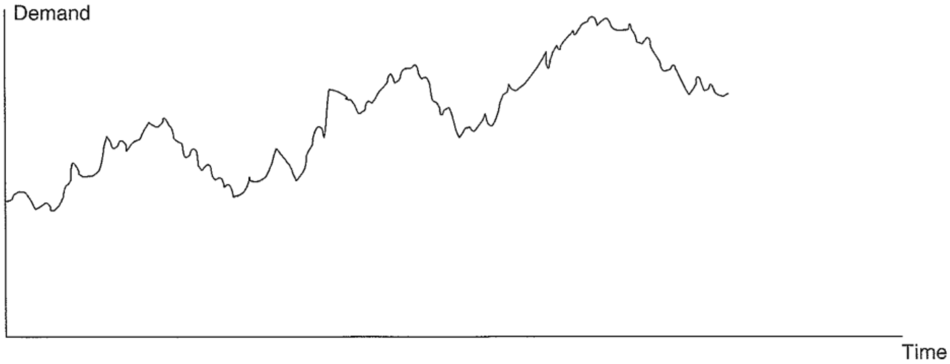

Figure 2.5: Example of time series containing all of the elements.

Moving average

Stevenson [22] mentions that forecasting using lag of only one period (i.e. naive method) has the weakness in the fact that no smoothing happens whatsoever. This can be overcomed by expanding the amount of historical data a forecast is based on. A moving average forecast estimates the demand for the next period as the mean of demand in the past n periods. It can be computed using the following equation:

$$F_t = MA_n = \frac{\sum_{i=1}^{n} F_{t-i}}{n}$$

F_t - Forecast for time period t

MA_n - n period moving average

A_t - Actual value in period i

n - Number of periods (data points) in the moving average

For example, MA_3 would refer to a three-period moving average forecast, and MA_5 would refer to a five-period moving average forecast.

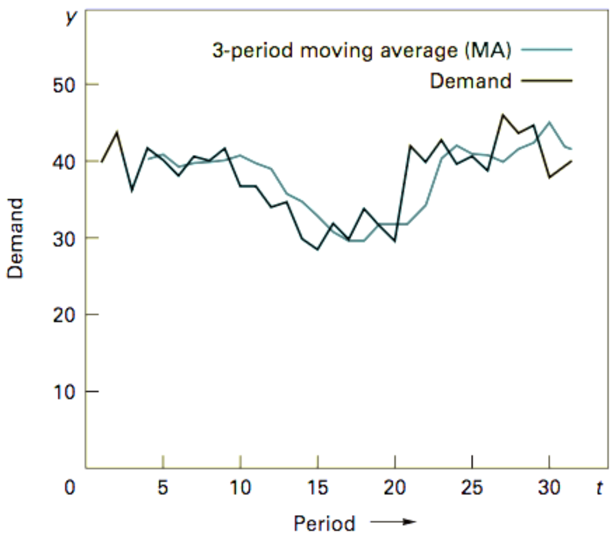

Figure 2.6: A moving average forecast tends to smooth and lag changes in the data[22].

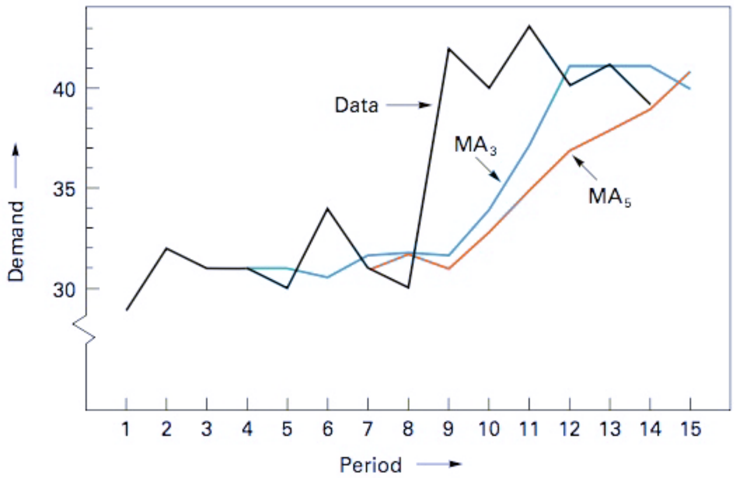

Figure 2.7: The more periods in a moving average, the greater the forecast will lag changes in the data[22].

Exponential smoothing

Exponential smoothing is a forecasting method where each new prediction bases itself on the previous one plus a percentage of the difference between that forecast and the actual value at that time period.[22] In other words, it produces forecasts which can be seen as weighted averages of past observations. Going into the past, these weights decline exponentially with each period. The more recent the observation, the higher the weight. Mathematical formula is as follows:

$$F_t = F_{t-1} + \alpha(A_{t-1} - F_{t-1})$$

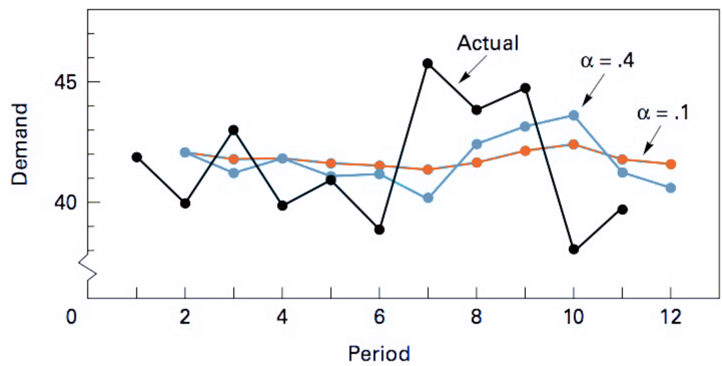

Figure 2.8: The closer coefficient is to zero, the greater the smoothing[22].

Second order exponential smoothing

If a time series shows signs of a linear trend presence, an adequate forecasting method might be trend-adjusted exponential smoothing or second order exponential smoothing. In this model, level and trend factors are successively modified, based on the latest observed demand in the following manner [23]:

$$F_{t+1} = L_t + T_t$$

$$L_t = \alpha A_t + (1-\alpha)(L_{t-1} + T_{t-1}) = \alpha A_t + (1-\alpha)F_t$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$$

Holt-Winter's method

When seasonality is present, often used method is Holt-Winters seasonal method. In addition to level L_t and trend T_t equations, this method uses also seasonal component S_t. Smoothing parameters are noted as α, β and γ. Another variable required by this method is m, which describes the period of seasonality. [19]

$$F_t = L_T + T_t + S_{t-m}$$

$$L_t = \alpha A_t + (1-\alpha)(L_{t-1} + T_{t-1}) = \alpha A_t + (1-\alpha)F_t$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$$

$$S_t = \gamma(F_t - L_{t-1} - T_{t-1}) + (1-\gamma)S_{t-m}$$

2.3 Inventory management

Inventory is any store of goods. Which kind of items are stored is dependent mostly on the industry. In case of manufacturing companies, these are mostly supplies of finished and semi-finished products, purchased goods and raw materials. Spare part, tools, and other supplies are also commonly stored.

According to Stevenson [22] the most important functions of inventory are:

1. **To meet anticipated customer demand.** Also called anticipation stocks because they are held to satisfy expected (i.e., average) demand.

2. **To smooth production requirements.** Here belong the seasonal inventories which are build up during preseason periods to meet overly high requirements during seasonal periods.

3. **To decouple operations.** Inventories are often used as buffers in order to maintain continuity of production so it wouldn't be disrupted by the breakdowns or any other events. They permit operations to continue while the problem is being resolved. Disruptions in deliveries from suppliers are also often covered by buffers.

4. **To protect against stockouts.** The risk of lack of materials or products can be reduced by the presence of safety stocks.

5. **To take advantage of order cycles.** Very often it is much more economical to purchase or produce in certain quantities. This can even make sense if these quantities exceed immediate requirements. Inventories allow companies to buy and produce in economic lot sizes and store these goods for later use. This results in periodic orders or order cycles.

6. **To hedge against price increases.** Similarly to the previously mentioned function inventories can also be used to stock goods which were more economical to buy earlier than the time they are required.

7. **To permit operations.** Production operations take time and are not instantaneous. Thus there is always some work-in-process inventory. These inventories are referred to as pipeline inventories.

8. **To take advantage of quantity discounts.** When suppliers give discounts on large orders.

Chapman [3] states that there are two basic categories of inventory replenishment models:

2.3.1 Quantity-based inventory models

Quantity-based (continuous review) inventory models basic assumption is that inventory is continuously monitored, so the control system can at any time tell exactly what the inventory position is. The figure (reference) is unrealistic since an assumption is made that the inventory is replenished immediately when required. In reality, this is not the case, and the time the time it takes to make that replenishment is called the lead time. Taking lead time into effect is achieved by ordering the goods when the reorder point is reached. This is the quantity that should be enough to cover demand during the time it takes to replenish the inventory. It is calculated as follows:

$$R = A * L,$$

where R is a reorder point, A is the average daily demand and L is lead time.

Demand is not necessarily constant from day to day and often some unexpected disruptions appear. To make sure that there is sufficient inventory on hand to meet the demand there should be some kind of buffer. This buffer is referred to as safety stock. According to Chapman [3] the amount of safety stock is generally dependent on two issues: T

- The variability in the demand
- The expected probability of not hitting a stock-out, also referred to as a service level.

When a normal distribution of demand is assumed, safety stock formula is typically given as:

$$S = z\sigma$$

,

where S is safety stock, z is the statistical z score corresponding to the desired service level, and σ is the standard deviation. By looking at the formula it is easy to conclude the larger the service level also means larger the safety stock for any given variability in demand.

Combined with the safety stock, reorder point formula becomes:

$$R = A * L + z\sigma.$$

2.3.2 Time-based inventory models

Time-based inventory models allow the inventory to be used without keeping records updated until a certain time has elapsed. The procedure is to count the inventory and then determine the suitable replenishment quantity while taking the lead time into consideration.

Due to the fact that the actual demand could greatly exceed the predicted one this model is being used less frequently in practice. Another reason for this is that using quantity-based inventory model is nowadays much simpler and less expensive due to existence of the computer-based systems.

2.3.3 Basic inventory policies

Quantity-based and Time-based inventory models can be further split up depending weather order quantity and order period are constant or variable. There are four most commonly used inventory policies which can be easily represented using simple matrix. This matrix is shown in table 2.1.

Order quantity \ Time	Constant order intervals	Variable order point
Constant	(t,q)-policy	(s,q)-policy
Variable	(t,S)-policy	(s,S)-policy

t - Interval between two orders
q - Fixed(economic)order quantity
s - Reorder level
S - Difference between current and target inventory level

Table 2.1: Basic inventory models.

To better understand policies mentioned in the table 2.1, let us provide a brief description of each of them:

- **(t,q)** - Order up to a level S, every time the inventory level drops below the Reorder Level s. Often referred to as a Min-max system.

- **(s,q)** - This policy means ordering Economic Order Quantity q, every time the inventory level drops below the Reorder Level s. It is also called a two bin system as it can be viewed as having two bins: One is the Cycle Inventory, and another bin consists of Demand during lead time and Safety Stock.

- **(t,S)** - Order up to a level S, every t time periods.
- **(s,S)** - Order up to a level S if Inventory Position is less than or equal to Reorder Level s.

Clearly, some of these policies require more effort and resources than other. However, they are often more accurate and more reliable. To cope with this, companies often decide to use different inventory policies on different parts so that their efforts can be allocated in the best possible way.

2.3.4 Classifiction system

Inventory management also pays attention to the importance of parts regarding profit potential, dollars invested, sales or usage volume, or stockout penalties. It is very reasonable to devote a different amount of attention to different items according to their relative importance in inventory. The A-B-C approach classifies items according to some measure of importance, usually dollar value per unit multiplied by usage rate in some time period. The idea to focus effort where it counts the most. Typically, three classes of items are used: A (very important), B (moderately important), and C (least important). Inventory control efforts are then allocated accordingly. [22] The "rule of thumb" for separating A, B and C items that is often used is to first list all items in order of highest relative importance to lowest. The top 20% of the items often represent the A items, those between 20 and 50% are the B items, and the lowest 50% are the C items. [3]

2.4 Correlation analysis

2.4.1 Exploring the correlation

Important step before any statistical analysis of the two variables is creating a scatter plot. Any relationship that exists between them becomes easily visible.

"To distinguish it from other graphic forms, we define a scatterplot as plot of two variables, x and y, measured independently to produce bivariate pairs (x_i, yi), and displayed as in-dividual points on a coordinate grid typically defined by horizontal and vertical axes, wherethere is no necessary functional relation between x and y." [7]

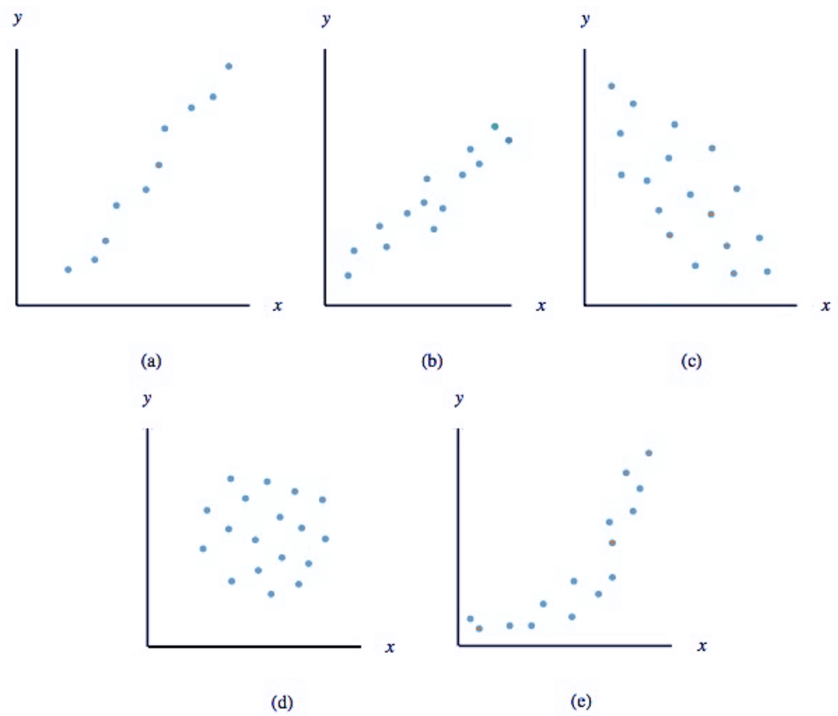

Figure 2.9: Scatterplots illustrating various types of relationships: (a) positive linear relationship; (b) another positive linear relationship; (c) negative linear relationship; (d) no relationship; (e) curved relationship. (introtostats)

It is possible to make a good visual impression of the strength of the relationship between two numerical variables by looking at a scatterplot. However, making more precise statements requires representing this relationship numerically. To do this, we use a correlation coefficient, and most commonly Pearson's sample correlation coefficient. It is a dimensionless index used as a measure of the strength of linear relationship between two variables. Its values range between 1 and 1, where 1 and -1 mean total positive and total negative linear correlation and 0 represents no linear correlation. Its calculation is based on z scores and given by:

$$r = \frac{\sum Z_x Z_y}{n-1}$$

Pearson's sample correlation coefficient is often recommended because its mathematical and statistical properties have been studied in detail. [15]

2.4.2 Principle of least squares

The objective of regression analysis is to use information about one variable, x, to predict the value of a second variable, y. These two variables have different roles in regression analysis :

- **Dependent variable** denoted by y. This is the variable whose value is to be predicted. The dependent variable is also sometimes referred to as response variable.
- **Independent variable** denoted by x. This is the variable used to predict the dependent one. The independent variable is also referred to as the explanatory variable, predictor variable or input variable.

Sometimes, these two variables are connected by an exact straight line relationship. This means the scatterplot will exhibit a linear pattern. In these situations, a straight line can be valuable in summarizing the observed dependence of one variable on another.

An equation of a line is

$$y = a + bx,$$

where a is referred to as the intercept of the line. It is the height of the line above the value $x = 0$. Slope of the line b is the amount by which y increases when x increases by 1 unit. [15]

Many methods are available for obtaining estimates of parameters of a line. The method discussed here is the principle of least squares. In this method, the criterion function for obtaining estimators of parameters a and b is based on the vertical deviations which show how good a particular line describes the relationship between variables. Vertical deviations represent the vertical distances between the actual y values and the fitted line. A point that lies above the line results in a positive vertical deviation and a point that lies below the line in a negative vertical deviation. A line is considered a good fit to the data if the deviations from it are small in magnitude. Parameter estimates are chosen to minimize the sum of squared deviations.

$$\sum[y-(a+bx)]^2 = [y_1-(a+bx_1)]^2 + [y_2-(a+bx_2)]^2 + ... + [y_n-(a+bx_n)]^2$$

The line that minimizes this sum is called least-squares line or the sample regression line:

$$\hat{y} = a + bx,$$

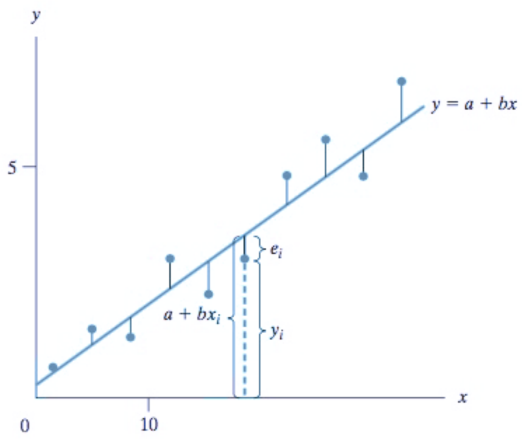

Figure 2.10: Diagram for least squares criterion showing the vertical deviations whose sum of squares is minimized.(Miller and freund)

where $\hat{y}$ the prediction of y that results from substituting a particular x value into the equation.

The slope of the least-squares line is:

$$b = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2}$$

and the y intercept is

$$a = \bar{y} - b\bar{x}$$

A prediction of a value is obtained by substituting x into the least-squares line. Then e_i, the error in predicting the value of y corresponding to the given x_i, is

$$e_i = y_i - \hat{y}_i$$

3 Case study introduction

3.1 Company

ALNO AG is a kitchen manufacturing company located in Germany. Its business area involves the development, production, and sale of kitchen furniture globally. From its four national and international sites, the company supplies markets all across the world. In addition to the core ALNO brand, the ALNO Group is also made up of the PINO, WELLMANN, PIATTI and Foster Swiss Steel Kitchens or ALNOINOX brands. Through these brands, the company offers a full range of kitchens; from low-price to premium price. The ALNO Group operates in more than 64 countries worldwide. In the financial year 2015, it generated sales in the amount of 522 million euros. [1]

Headquarters of ALNO Group is located in Pfullendorf, Germany. Here the kitchen brands Alno, Tielsa, and Piatti are produced. This location also produces base covers, carcase elements and crown moldings for the other production facilities within the Group.

This Pfullendorf location includes four plants:

- **Plant 1** - Serial production and make to order front manufacturing
- **Plant 2** - Make to order body elements
- **Plant 3** - Assembling and shipping
- **Plant 4** - Production of glass and ceramic surface parts and painting

ALNO owns two more plants in Germany; one in Coswig at which kitchen brand PINO assembled and the other one in Enger where Wellman is made. Arbon in Switzerland is the location where Steel kitchens (ALNOINOX) are manufactured.

3.2 Product

A journey of every kitchen begins with kitchen planners. These employees are responsible for direct communication with the customer, design of the kitchen and input of an order into the system. The latter two are done using the so-called COE software. Once the order has been received and loaded into COE, it is first split up into elementary units. These units are kitchen cabinets, and they represent the primary product of ALNO. Therefore, kitchens are always viewed as the group of cabinets.

When making an order, a customer has virtually unlimited design possibilities. From glass or ceramic front panels to painted elements: the options for combining the ALNO range of colors are unlimited. A wide range of customisation possibilities is available in a way that virtually anything can be adapted to personal wishes.

The most common distinctions between kitchens, however, are the visible parts, such as fronts or working plates. Elements of the cabinet body, such as side or back panels, vary mostly only in dimensions which are standardized. This large amount of variants makes the production process very challenging to plan and organize. As a result, all of the subassemblies are split into three categories:

- **Parts produced to order**
  Also called lot size one parts by employees. Here belong those parts which make the kitchen unique, mostly fronts and working plates. Part of this group are also carcase elements with nonstandard dimensions since they are not produced on stock. Part of the company in charge of manufacturing these parts follows Make-to-Order (MTO) logic of planning.

- **Parts with no relation to customer**
  These parts are manufactured in a plant 1, and are often referred to as serial production parts or Make-to-Stock (MTS) parts. Here belong the carcase elements of standard dimensions and some simple fronts. They are produced on stock and only sent to an assembly when needed for a final product. In other worts, they don't require customer order to be produced.

- **Purchased parts**
  Not all kitchen components are produced within ALNO facility. Many non-wooden parts like hinges, handles, sinks and various appliances are bought from external suppliers. Here also belong some fronts and carcase elements which cannot be manufactured by the company due to lack of technology. Depending on their usage rate, some of these parts are held on stock while others are purchased only when customer order is placed.

This classification helps with the efficiency and the flexibility of the production process. Finished ALNO cabinets are only assembled once an order is placed and confirmed. Therefore, reffering back to the chapter **??**, it can be said that planning of ALNO cabinets as final products belongs to Assemble-to-Order (ATO) category.

3.3 Serial production

ALNO serial production is located in the plant 1. As mentioned in chapter 3.2, it is responsible for the manufacturing of subassemblies which are produced on stock. Serial production parts are mostly carcase (body) elements such as sides and back panels, shelves, toe kicks, drawer segments, etc. Main raw materials used to manufacutre these parts are refined chipboards and edge bandings.

The serial production process begins after SAP generated production order has been confirmed and released as described in chapter 3.4. The step which follows is the optimisation of the board cutting plan. This is done with the help of specialized software which helps the worker to define an optimal way of cutting panels out of a raw chipboard. The optimal cutting plan is the one which creates the least waste and very often it means producing a few panels more. After the plan is confirmed, the number of produced panels is reported back to the SAP so that proposed quantity can be replaced with the realized one.

This information is also loaded into a computer which controls the saw machine. It cuts the raw panels according to the layout and allocates them further into one of the so-called AK machines. The functions of the AK machines are following:

- **Formatting**
  Saw is sometimes not able to cut panels to exact needed dimensions. In that case, additional, more precise cutting is required.

- **Edging**
  The technology used is either laser or polyurethane + heat. Edging can be 2-sided or 4-sided.

- **Drilling**

- **Doweling**
  Adding dowel pins to the panels for the joining purpose.

Finished serial production parts split into two ways. Most of them are sent via a conveyor belt to the high rack warehouse. These parts are later forwarded to the assembly plant to be used for ALNO and Piatti cabinets. Other group of Pparts is shipped directly as subassemblies to the daughter companies Pino and Wellman.

3.4 Current planning

3.4.1 SAP planning

Serial production materials are planned using SAP PP Material requirements planning. Its primary task is to guarantee material availability in the final assembly, that is, to create timely procurement proposal for production with regard to requirement quantities. MRP types are available within SAP, but in the case of ALNO type PD is mostly used. It refers to deterministic planning and implies that the MRP creates order proposals so that the exact quantity of the requirements is matched. When creating the order proposal, it has to adhere to the lot-sizing procedure set in the material master MRP 1 view.

This process takes into consideration two consumption logic parameters - the minimum lot size and the rounding value, both of which are set manually by the MRP controllers. The rounding value is defined as the quantity of the material able to fit onto one pallet and is important for the efficiency of the internal transport. According to the example from the SAP documentation[21], if the Rounding Value is 20 units, the order quantity 70 units, and the minimum lot size 15 units, the system will always round up the planned order to a multiple of 20 units. The idea behind minimum lot size is ensuring the production in economic quantities.

When calculating shortage quantities i.e. quantities to be produced SAP takes into account two types of planned independent requirements:

- **VSFB** represent the actual requirements placed on a specific day. These are derived from the demand for end products and are transferred to SAP automatically after BOM explosion.
- **ZSFB** requirements are calculated as the mean of VSFB type requirements for the present-day and the next ten days. Therefore, ZSFB first appears on the tenth day after the present-day. Any existing VSFB requirements on that day are then subtracted from this number. Let t be the day and $t \geq 10$, then ZSFB requirements can are calculated as follows:

$$ZSFB_t = \frac{\sum_{t-10}^{t} VSFB_i}{11} - VSFB_t.$$

These ZSFB requirements then represent a forecasted demand for a part.

3.4.2 Role of the MRP controllers

MRP controller plays a key role in the planning process. He is responsible for material requirements planning accuracy and material availability. His task is to make sure that the materials required to manufacture finished products are available on time. He evaluates the MRP generated order proposals and has a possibility to adjust their quantities manually if necessary. These manually adjusted quantities should, however, also take into account previously mentioned consumption logic parameters.

According to SAP documentation [21], MRP controller can use key figures such as range of coverage and exception messages to help them complete their tasks. His responsibilities include:

- Executing and monitoring material requirements planning
- Converting order proposals
- Monitoring stocks and range of coverage

In addition to this, controllers occasionally use the possibility of manual adjustment of MRP generated recommendations. They do this to counteract faulty estimations of the requirements made by the system. This matter will be discussed in more detail in the upcoming chapters.

3.5 Problem definition

Customization strategy is becoming more and more pursued by companies across many industries to gain competitive advantage. A better match between customer requirements and product offering is expected to increase customer satisfaction, and organizations must obtain that goal in a way that makes them money.[16]

Studied company ALNO AG belongs to one such industry. To stay competitive in kitchen manufacturing business, it is important to allow your customers plenty of customisation possibilities. This, of course, has its tradeoffs. More customisation usually means increasingly complex manufacturing process and requires very sophisticated production planning and scheduling to keep the costs at a reasonable level. This need for reliable and accurate plans and better forecasting arises as a result of more uncertainty in demand. Of course, these fluctuations can easily be covered by having large inventories, but this is un-optimal and leads to unnecessary costs. Moon et al. (1998:44) state: “Inventory exists to provide a buffer for inaccurate forecasts, thus the more accurate the forecasts, the less inventory that needs to be carried, with all the well-understood cost savings that brings”.

SAP has been operating in ALNO for five years, and since then it has collected a large amount of data. However, this data was never thoroughly analyzed, and there were no attempts to obtain useful knowledge from it. ALNO representatives are interested to see what kind of value can be extracted from this data. Primarily, they pursue improvements in the field of demand forecasting and inventory management. They believe that more insight into the available data will allow the company to make better decisions and reduce inventory levels.

Production planning and scheduling in ALNO AG serial production is an extremely complex and time-consuming process done by many employees across different departments. It uses a wide range of inputs from several sources and doesn't have a precisely defined procedure. This complexity and lack of structure lead to mistakes and unnecessary costs. Solving these problems requires a study of the process together with a review of its inputs and decision-making variables.

3.6 Scope of the project

The largest potential for improvement lies in the optimizing the planning process of parts produced on stock. These parts are manufactured in large quantities and have much better data management compared to other components. What more, they should be the focus of forecasting process as customer orders do not trigger their production. As mentioned earlier in chapter **??** they are divided into two categories, and this study is going to cover only those parts which are used for ALNO and Piatti cabinets.

SAP database holds records for about 1500 individual parts which fit into this category. However, not all of them are considered relevant for this study. Due to the fact that there is a lot of variation in the amount of available historical data, some limitation on how much information it is considered sufficient had to be established. It was decided that only those parts whose data dates back to at least one year in the past should be found as relevant. Another limitation had to be put on the volume of the demand since it is very low in some parts. If a part is required extremely infrequently and in very low quantities, there is little that can be done in terms of improving statistical forecasting accuracy and inventory management process. Moreover, there is a concern that the data from these parts may significantly skew the results. Therefore, this study will cover only those parts which have average daily demand A larger than 1.

3.7 Methods

Both qualitative and quantitative methods will be employed with the goal of delivering the desired results. This combination can improve the research by ensuring that the any possible limitations of one approach is balanced by the strengths of another.

3.7.1 Quantitative methods

Interviews and observation serve as a way of acquiring qualitative data. They are the starting point of the research as they provide the ability to identify the issues of relevance. This is particularly important in circumstances where little is known about the case and improvement possibilities. These methods were mainly used in the part of the thesis which analyses the current planning process of the company. In addition to that, qualitative information also greatly contributed to the understanding of the case context and the situation in the company.

3.7.2 Qualitative methods

However, the main focus of this case study is the quantitative data. As mentioned before in the chapter 3.6 the data used in this thesis is mainly the historical demand. It includes a large amount of information making it almost impossible to process and analyze part by part. Hence, some sort of process automation was necessary. A decision was made that the central part of data collection, processing, and analysis should be done using computer algorithms. A specially designed computer program not only allows more accurate but also much faster processing of information compared to some other approaches. Because all manipulation of the data is done with code, it can be repeated an infinite number of times. In other words, once an algorithm is written for one dataset, it is very easy to reproduce it on many others. In addition to all this, algorithms developed during this study leave open the possibility of developing a software which could serve as a production planning support system. Such kind of assistance may prove to be of a great value to the company.

Python programming language was chosen as most suitable for this purpose. The reason for this is its built-in database functionality which makes manipulation with datasets very convenient. A variety of data persistence features and libraries for interfacing with Microsoft Excel were also important factors.

Algorithms were used to complete various tasks throughout this study. Some will be explained in the corresponding sections using pseudocode and textual explanation.

However, due to the length of the written code, most of them will only be mentioned and available to the company on request.

Some of the main tasks completed using python programming language are:

- Collection of data from the Excel spreadsheets
- Data processing and time series transformation
- Outlier identification and removal
- Classification of parts
- Implementation of forecasting methods
- Forecasting methods performance determination
- Inventory policy simulation
- Optimization of simulation parameters
- Development of numerous plots used throughout this study

Microsoft Excel, on the other hand, was used for tasks where automation was not necessary. This applies to correlation and regression analysis as well as creating some charts.

4 Case study research

4.1 Analysis of the current forecasting process

As mentioned before in chapter 3.4, SAP generated order proposals occasionally need to be adjusted because they do not match real requirements. One of the reasons behind these mistakes is a high dependence of material requirements planning efficiency upon accuracy of the bill of material (BOM). Analysis of the current planning process revealed that BOM is inaccurate for some products. In these cases BOM explosion reports wrong required quantities which in turn leads to a generation of inaccurate order proposals. Controller also adjust order proposals when they believe SAP forecasting based on ZSFB is wrong.

To estimate future requirements when BOM is not correct, or order proposals are not trustworthy for some other reason, MRP controller often relies on other guidelines available within SAP MD04 transaction.

4.1.1 AMC based prognosis

When required quantities derived from the existing orders in the system are wrong, that is when BOM is inaccurate, the only way to assess future demand is to use the historical data.

The way MRP controllers do this is by looking at the average material consumption of the previous months. From the interviews it was concluded that this is most commonly used guideline when manual adjustment is necessary. The MD04 table gives a possibility to see average material consumption for one, two, three, six and twelve calendar months in the past. It is important to mention that it is not possible to see the average consumption of the current month. Consequently, this can lead to faulty estimations in the case of sudden change in a required quantity of material. As an illustration of this issue let us consider example shown in the figure 4.1.

The graph in the figure 4.1 shows daily consumption of material 131222 for the months seven and eight of the year 2014. It can clearly be seen that there has been a sudden decrease in the daily requirements for month 8. They are considerably

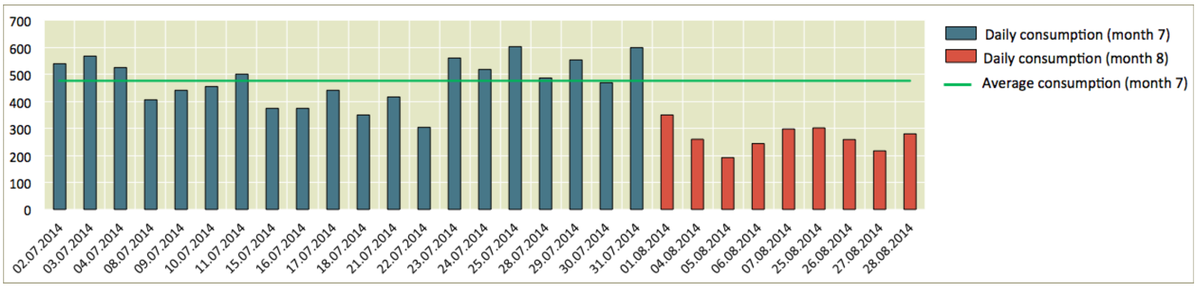

Figure 4.1: Average daily requirements of the part 131222 in a time period from 2.07.2014 to 28.08.2014.

lower than the average consumption for month 7 represented by the green line. In order to explain why this situation could lead to problems, let's suppose that in the middle of month eight MRP controller had to estimate future requirements for the purpose of creating production orders for the rest of the month. If he decided to use average monthly consumption of a previous month as a reference number, the estimation would equal 475. That would lead to large excess in production of the part considering that real requirements are much lower.

4.1.2 Bedvo based prognosis

Bedvo is a measure which ALNO has been using for a long time to predict the future requirements of a part. Unfortunately, as the variety of available cabinets became higher and the demand increasingly lower so did the reliability of this measure. Formulas required to calculate Bedvo number are given by,

$$BEDVO = \frac{R}{n},$$

$$n = \frac{N}{C},$$

where N is a number of cabinets currently on order, C is daily production capacity of cabinets, n is a number of days required to produce ordered cabinets by utilizing full capacity, and R is required quantity of material A within existing orders.

The objective of Bedvo number is the estimation of daily requirements for material A in the upcoming weeks. However, a thorough analysis of it revealed a few issues. The main downside is that it assumes the share of the material A within N orders to be constant. This may, however, not always be the case. Material A could, for example, be only a part of first half of cabinets currently on order. Distributing the requirements evenly across the whole period of N orders would then lead to the lack of material in the first and excess in the second half of the period. To put it briefly, Bedvo number reliability is proportional to a uniformity of materials demand.

4.1.3 Füllgrad based prognosis

"Füllgrad" is a rate of capacity utilization in the specific week. Therefore, when the "Füllgrad" is at 100%, this week is filled with orders, and it is no longer possible to add any more to it. This also means that the material requirements for this period will most likely remain unchanged. On the other hand, when the "Füllgrad" is less than 100%, the requirements for the week are uncertain and have to be estimated. Füllgrad based prognosis is calculated by,

$$P = \frac{R}{n} \cdot \frac{1}{F},$$

where N is a prediction of average daily requirements within a certain week, n is a number of working days within that week, R is a known required quantity of a material A within that week, and F is Füllgrad.

Similarly to *BEDVO* number, this guideline assumes that the share of a material within orders will remain constant. It expects the cabinets which are still to be ordered in a specific week to have the same requirements for a certain material as the already ordered cabinets in that week. Because the volume of orders has in the last few years become increasingly lower and there are fewer weeks with Füllgrad of 100%, it is now much harder to say how many of material A will be needed in the following weeks.

4.2 Data preparation

4.2.1 Data collection

Few different options were considered when gathering the data which represents serial production parts requirements. In the ideal case, it would be derived from the end product demand because this opens a wide range of possibilities for different regression analyses. Linking the data directly to the sales of the cabinets requires using BOM explosion of final products. Unfortunately, interviews with the IT department revealed that this is not possible due to the limitations of HOST software where BOM is stored.

The best alternative was to use the machine gathered inventory data available in SAP. The inventory system records inputs and outputs automatically and stores this information in the SAP database. One can assume that every time a serial production part is removed from the inventory it means it is required for the end product. Consequently, this movement can be considered as its demand.

Accessing the SAP database which holds this information is possible through MB5B transaction. This transaction can generate a report containing inventory movement records for a given part in a given period of time. This allows seeing which type of movement happened at which date and in what quantity. It also presents some basic information such as material number, description, and price. After material number and time span are chosen, and the transaction is requested, SAP creates the table with all the bookings in the inventory which refer to this particular material. A possibility of exporting this table to the excel spreadsheet was used as a primary input for the later developed algorithm.

Bewertungskreis 0001			
Material 20105			
Bezeichnung BO UE 90 WS			
Bestand/Wert zum 01.01.0000	373 ST	2.246,16 EUR	
Summe/Wert der Zugänge	13.169 ST	79.690,30 EUR	
Summe/Wert der Abgänge	13.437- ST	81.289,66- EUR	
Bestand/Wert zum 31.12.9999	105 ST	646,80 EUR	

LOrt	BwA	S	MatBeleg	Pos	Belegnr	Buch.dat.	Menge	BME	Betrag Hauswähr
23	201		4900001898	1		09.01.2012	-15	ST	-87,23
23	201		4900001899	1		09.01.2012	-8	ST	-46,53
23	201		4900001900	1		09.01.2012	-6	ST	-34,89
23	201		4900001901	1		09.01.2012	-13	ST	-75,6
23	201		4900009533	1		10.01.2012	-10	ST	-58,16
23	201		4900009534	1		10.01.2012	-5	ST	-29,08
23	201		4900015177	1		11.01.2012	-8	ST	-46,53
23	201		4900015178	1		11.01.2012	-6	ST	-34,89
23	521		4900018030	1		11.01.2012	206	ST	1.198,10
23	201		4900020121	1		12.01.2012	-4	ST	-23,26
23	201		4900020122	1		12.01.2012	-10	ST	-58,16
23	201		4900025297	1		13.01.2012	-9	ST	-52,34
23	201		4900025298	1		13.01.2012	-8	ST	-46,53
23	201		4900030666	1		16.01.2012	-7	ST	-40,71
23	201		4900030667	1		16.01.2012	-7	ST	-40,71
23	201		4900036070	1		17.01.2012	-8	ST	-46,53
23	201		4900036071	1		17.01.2012	-8	ST	-46,53
23	201		4900041438	1		18.01.2012	-8	ST	-46,53
23	201		4900041439	1		18.01.2012	-13	ST	-75,61
23	201		4900046614	1		19.01.2012	-1	ST	-5,82

Figure 4.2: MB5B transaction exported in excel

4.2.2 Data cleansing

To analyze this data, it was necessary to transform it into a proper time series. The fastest and most efficient way was to develop an algorithm which would perform this task. In order to consider dataset a discrete time series, it's variables must be occurring at distinct, separate points in time. As it can be seen from the figure 4.2, existing dataset often contains more than one individual entry per day. This means it does not fulfill requirements of a discrete time series. Hence, the first task of the algorithm is to add together all of the entries made on the same day.

If there were no entries on a particular day, then the demand on this day should be considered zero. Exceptions to this rule are non-working days. It would be wrong to consider holidays and weekends to be the days with zero demand. This could result in an appearance of irrelevant patterns during the data analysis process. To cope with this issue, an algorithm had to be provided with the company calendar which covers the whole period from the beginning of recorded data until the present day. This calendar was found in SAP and then converted to excel file for easier handling.

Date	Truth
01.01.2012	1
02.01.2012	1
03.01.2012	1
04.01.2012	1
05.01.2012	1
06.01.2012	1
07.01.2012	
08.01.2012	
09.01.2012	
10.01.2012	
11.01.2012	
12.01.2012	
13.01.2012	

Figure 4.3: Part of the excel calendar table.

Figure 4.3 shows excel table containing all the considered dates where truth value of 1 corresponds to the non-working days. The algorithm reads this spreadsheet and removes all of the irrelevant days from the time series.

Another step in data gathering process was to remove irrelevant bookings. The simplest way of identifying these was to look at which movement types were assigned to them. Some of the bookings represent a correction of inventory records and as such should not be analyzed. These corrections are required when inventory level in the system needs to be brought into an agreement with the findings of the physical inventory or after a booking mistake is made. In consultation with employees, it was decided that only bookings with movement types 201 and 953 should be identified as relevant demand. Procedure in the algorithm neglects all other bookings.

It is important to mention that all of the available material data within MB5B transaction such as description, initial inventory, material number, and price are stored into corresponding variables.

```
Algorithm 1: Transforming a data into time series
Data: Entries from the excel report
Result: Time series
excelframe← dataframe holding inventory records read from excel
  excelframe[x][y] - column x, row y record in excelframe
for i = 1 to length(excelframe) do
    if excelframe[date][i] != 0 in calendar then
        next i
    end
    if excelframe[movement type][i] != 901 or excelframe[movement type][i] != 241
      then
        next i
    end
    if excelframe[quantity][i]> 0) then
        if there is a row in demand with same date as excelframe[i] then
            add excelframe[i] and existing record together in demand
        else
            new row in demand = excelframe[i]
        end
    else
        new row in demand = 0
    end
    demand ← time series
end
```

4.2.3 Outlier removal

An observation of the available data revealed a frequent appearance of values which differ from the expectation by a considerable degree. This is the case both with the values which are too high as well as too low. Such observations referred to as outliers. In this part of the thesis an appropriate strategy for dealing with outliers is developed.

Not only is this important for the relevance of this thesis results but can also improve any future research on the same data sets.

Hawkins [8] definition of an outlier is as follows: "An outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism.". Most often it is up to the analyst's to decide what will be considered anomalous. The main reason for identifying outliers is the fact that extremities may significantly skew a dataset and influence results of an analysis.

First, it is necessary to clarify some possible sources of outliers within the data. Companies project business described in the chapter **??** is considered to be the main reason behind values which are too high when compared to the rest of the dataset entries. Large projects often mean that the same kitchen will be required in large quantities. This further leads to a significant rise in demand for one cabinet program which results in an increase in requirements for parts included in this program. Projects are very stochastic events, can't be considered a rule and as such should be neglected in the data. Most often such large requirements for that particular cabinet type never appear again.

The second important source of outliers are the mistakes employees make while creating inventory records. While part of the booking process is automated, there is still a lot of manual scanning involved. The presence of human factor is a primary cause of mistakes in many processes, and there is a substantial probability that recording inventory movement is not an exception.

Lower outliers are also detected within the data. For example, if the requirement for a part averages at 300 units per day with a small variance it is very unlikely that it will drop to a value of zero or close to zero on a single day and increase again after that.

Detecting outliers requires structured approach. Many methods exist out there, but which is the most adequate depends mostly on the properties of the distribution. Engineering and statistics handbook [14] recommends generating a probability plot of the data before applying an outlier test. This is because most of the outlier removing methods assume normality. If we use them without any knowledge about the distribution, they could detect some extreme values which may in fact not be an outlier, but a result of non-normality of the data. Therefore, a plot combining data histogram and kernel density estimation was created.

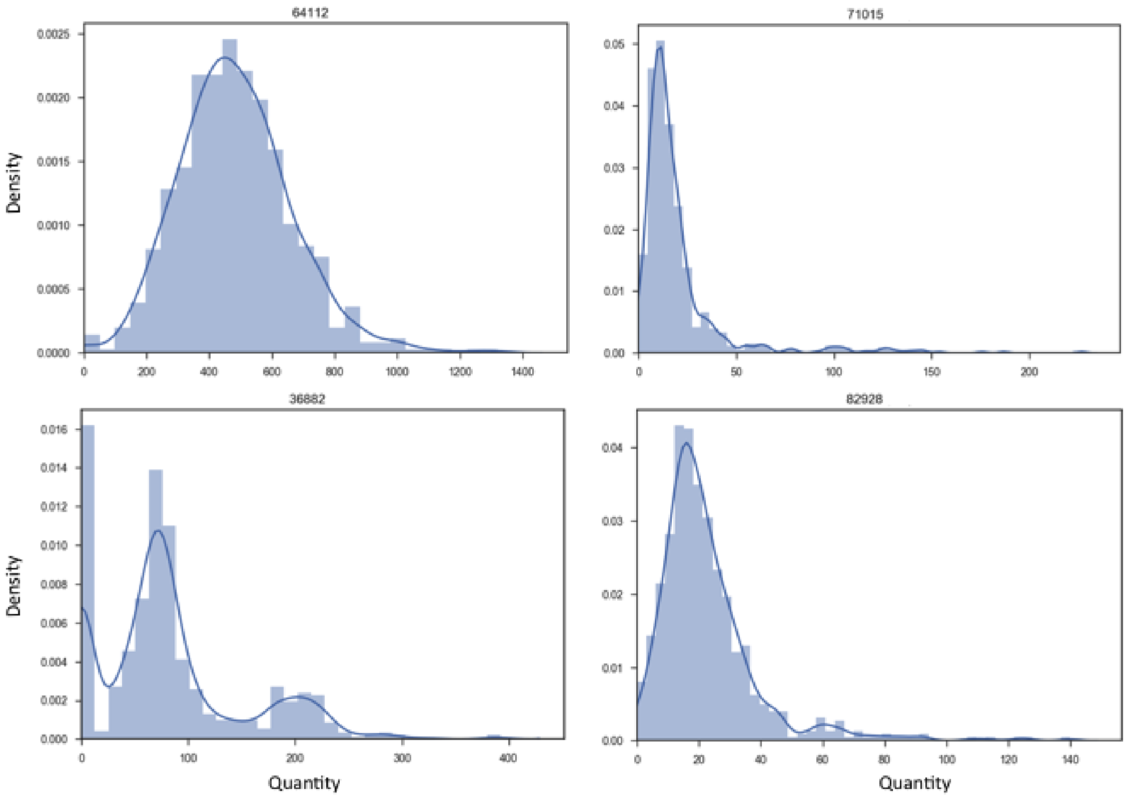

Figure 4.4: Histogram and kernel density estimation describing demand data of different parts.

The examples presented in the figure 4.4 show that datasets vary greatly in their histograms. Some of the distributions show normality, but there is also a lot of skewed and heavily-tailed ones.

M.Hubert and Vandervieren [12] in their paper propose an outlier identification rule for these kind of distributions based on an adjusted boxplot. The main idea behid it is the adjustment of a standard boxplot whiskers by the degree determined using the function of medcouple - a robust measure of skewness. Hubert and Vandervieren found this function through a series of simulations ran on a large amount of moderately skewed distributions.

In the case of normal distribution, the easiest and most commonly used method is an application of the three sigma rule. This rule states that an event is considered to be an outlier if it lies in the region of values larger than three times the standard deviation [13]. A second conventional method for outlier identification is a standard box plot. Both of these methods rely on the elliptical symmetry.

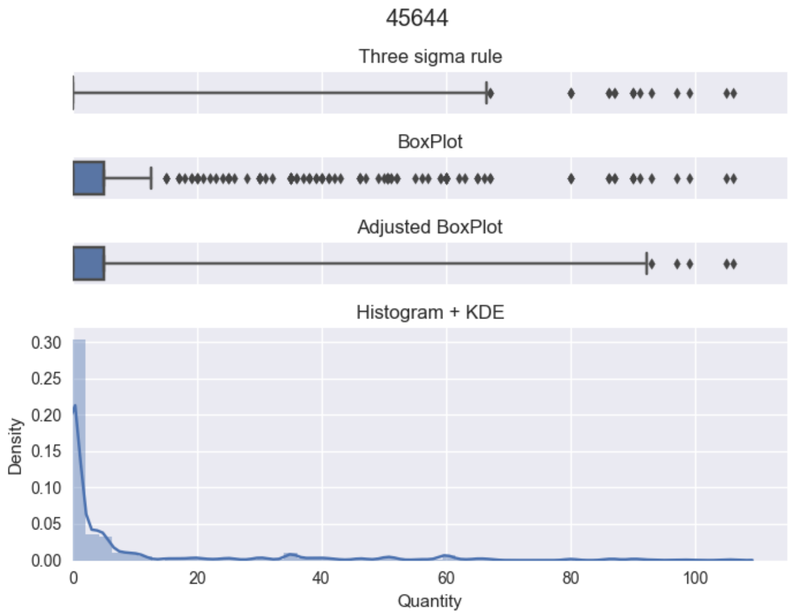

Figure 4.5: Comparison of different outlier removing methods performed on the demand data of the part 45644

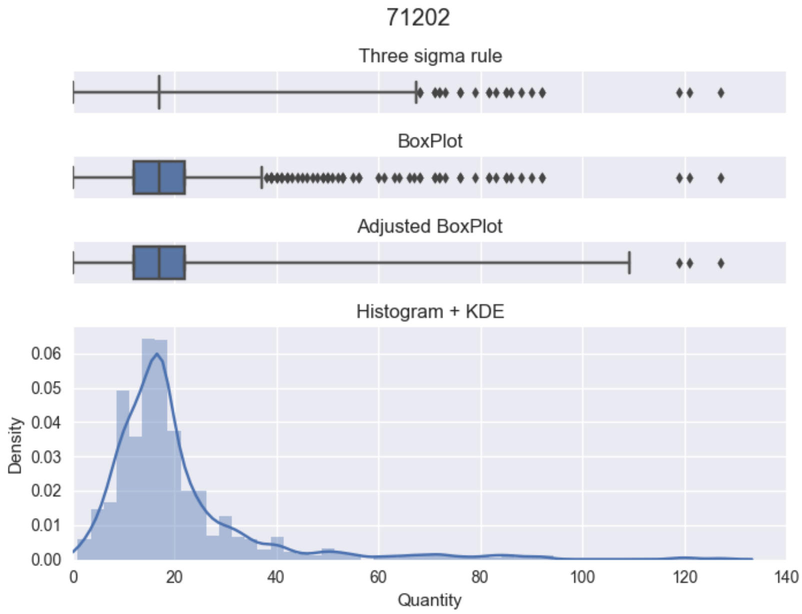

Figure 4.6: Comparison of different outlier removing methods performed on the demand data of the part 71202

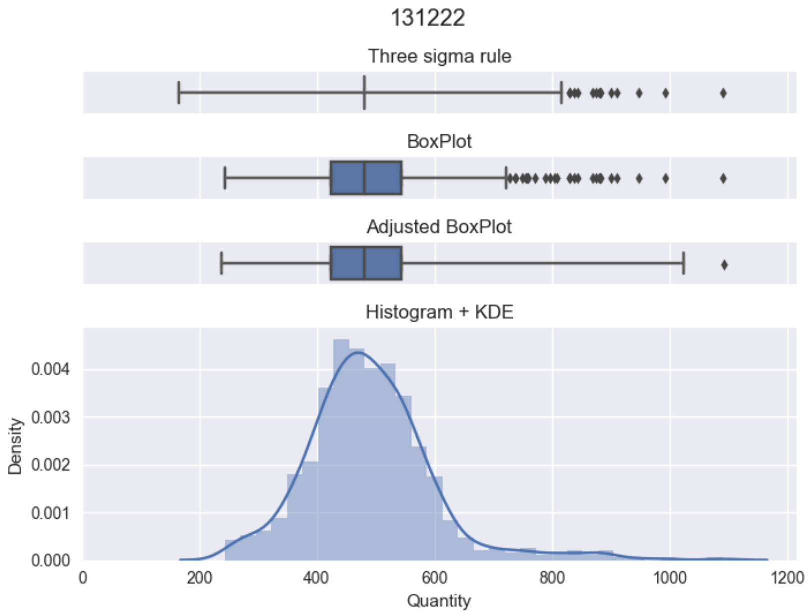

Figure 4.7: Comparison of different outlier removing methods performed on the demand data of the part 131222

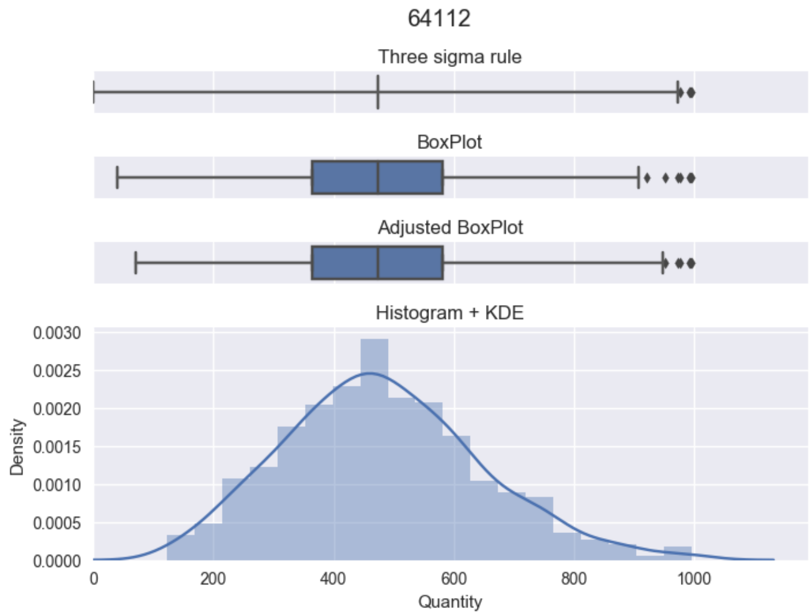

Figure 4.8: Comparison of different outlier removing methods performed on the demand data of the part 64112

All of the three mentioned methods were included in the algorithm to determine which one performs the best. Plots were developed in order to compare how these methods behave in different distribution scenarios.

In the figure 4.5 and 4.6 it can be seen how they respond to the skewed and heavily tailed distributions. Considering the shape of the distribution, it can clearly be concluded that three sigma rule and regular box plot method remove too many data points. Adjusted BoxPlot, on the other hand, recognizes that the distribution is skewed and adjusts the upper tolerance limit accordingly.

Distribution in the figure 4.7 is less skewed, but its right tail is still longer than left. First two outlier removing methods are symmetrical and set the limits on both sides to be at equal distance from the mean which is wrong in these situations. Here, modification of adjusted box plot whiskers is less significant, but still exists.

Demands for some parts follow the normal distribution, such as the one shown in the figure 4.8. In this case, three sigma rule and regular boxplot performances are satisfactory. Adjusted box plot behaves very similarly to these two methods since the adjustment of whiskers is insignificant.

Yet, despite all of this, due to the extreme diversity in demand distributions of the serial production parts, some modifications had to be made to the adjusted box plot. One of the problems occurred when the lower limit was larger than zero for the parts with intermittent demand. This is essentially wrong because the value of zero cannot be considered an outlier in the case of intermittent demand.

Another problem was that for some unusual distributions adjusted box plot method produced too many outliers. Because of this, some cases are considered special. A threshold value for the number of outliers was set at 2%, meaning that datasets with more than 2% of the entries identified as outliers are treated differently by the algorithm. In this case it creates a new dataset composed of detected outliers and considers only the ones further than two standard deviations from the mean as real outliers.

Algorithm 4 provides a more detailed explanation of the logic behind this outlier removing method.

Algorithm 2: Removing outliers from the time series

Data: Time series data
Result: Time series data without outliers
$l \leftarrow$ lower limit
$u \leftarrow$ upper limit
$uarray \leftarrow$ array of upper outliers
$larray \leftarrow$ array of lower outliers
$urep \leftarrow$ replacement value for upper outliers
$lrep \leftarrow$ replacement value for lower outliers
$Q_1 \leftarrow$ first quartile(25th percentile)
$Q_3 \leftarrow$ third quartile(75th percentile)
$MC \leftarrow$ medcouple
$IQR \leftarrow$ interquartile range
for *each demand* **do**
    if *number of zero periods in demand* $>$ *0.3 * length(demand)* **then**
        consider intermittent demand;
        l = - 1;
    else
        $l = Q_1$ - 1.5 * e^ (-1.25 * MC)* IQR
    end
    $u = Q_3$ + 1.5 * e^ (-4.25 * MC)* IQR
    $larray$ = time series values $< l$
    $uarray$ = time series values $> u$
    if *length(uarray)* $>$ *0.2 * length(demand)* **then**
        $urep$ = mean($uarray$)
        u = mean($uarray$) + 2 * std($uarray$)
        $uarray$ = *demand* values $> u$
    else
        $urep$ = mean(*demand* values)>0)
    end
    in *demand* find values = $uarray$ and replace with $urep$
    in *demand* find values = $larray$ and replace with $lrep$
end

4.3 Classification

In every company only so many resources are available, and where to use them is a matter of priority. In case of production planning and scheduling these resources are most often manpower, capital fund, and information technology. Since these resources are limited, in order to minimize costs and make the process as efficient as possible it is important to allocate them properly.

At first glance, it is very obvious that not all serial production parts in ALNO are equal. As mentioned earlier in the chapter 4.2.3 the distributions of their demands are very diverse. Some of them show very stable behavior while others seem quite unpredictable. Their average daily demand ranges from below one to hundreds of pieces a day. Furthermore, they differ in their prices.

Considering the number of parts in serial production and the fact that it is virtually impossible for each of them to be treated differently, the need for their classification arose. After taking a closer look at some of the main characteristics, three were selected as most important for the production planning and inventory management: the price of a part, the volume of its demand, and the demand volatility. These characteristics were represented using adequate descriptive statistical measures. The requirement was that they are be simple to calculate and understand so that classification of a new part could be done without the use of complex algorithms.

However, creating individual dimension for each of the mentioned characteristics would result in too many groups of parts. For the sake of simplicity, two of these characteristics were combined into one dimension. After thorough analysis, a decision was made to reduce the number of dimensions to the following two:

- **Relative importance of the part**
  It combines the monetary value of the part and the volume of its demand. Let I represent the relative importance of the part, then:

$$I = A * P,$$

$$A = \frac{\sum_{i=1}^{n} X_i}{n}.$$

  A - average daily demand
  P - price of a part
  X_i - demand in period i
  n - number of periods

- **Variability of parts demand**
  It is represented using coefficient of variation C_v. This measure is expressed as a percentage and calculated as follows:

$$C_v = \frac{\sigma}{A} * 100\%,$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(X_i - A)^2}{n-1}},$$

  where σ represents standard deviation.

According to relative importance the parts are divided into:

- **A parts** are the most profitable items. They require high level of attention and it is important that MRP controllers are informed in realtime about any exceptional situations during their production because avoiding stockouts is priority. Their inventory control should be tight and reorders frequent.

- **B parts** are parts of medium importance and they can be considered average. Nevetheless it is still important to put sufficient efforts into their planning and to carefully monitor them.

- **C parts** are of a low value, volume or both. Investing a lot of resources in their planning is therefore not justified. They should be reordered less frequently then parts from previous classes. In theory, stockouts in case of C parts are not very dangerous for the company.

According to demand variability the parts are divided into:

- **X parts** should be viewed as easy to forecast as the demand for them fluctuates slightly around a constant level. These items are characterized by relatively constant and non-changing usage over time.

- **Y parts** usage is neither constant nor irregular. In their case, it's harder to obtain an accurate forecast. Their forecast can improve if there is a presence of trends.

- **Z parts** parts are very unpredictable. Their usage fluctuates strongly or occurs intermittently which means that forecasting them is almost impossible. Some of these parts have periods with no consumption at all and are considered parts with intermittent demand.

Splitting items into A, B, C, X, Y and Z classes is relatively arbitrary. As a support when determining how boundaries for different classes should be set, one should use the visualization of dimension distribution. Figure 4.10 and 4.11 shows these distributions. This approach to segmentation was chosen because it is simple, a good fit for the distributions and creates groups of adequate sizes.

Python algorithm used for this classification determines the borders of classes automatically. It first calculates the coefficient of variance and relative importance for each part and then creates two separate datasets where the parts are sorted according to these two values.

First 20% of the parts in the dataset containing information about coefficient of variance C_v are assigned to class X, following 20% to the class Y and the rest to the class Z. The same procedure is performed on the relative importance dataset. The algorithm then compares these two datasets and uses intersection operation to form groups. As a result of this procces, parts are split up into 9 groups in a way represented by the figure 4.9. Algorithm 6 gives a more detailed explanation of the procedure. Table 4.1 shows how many parts are assigned to each group after this classification is performed.

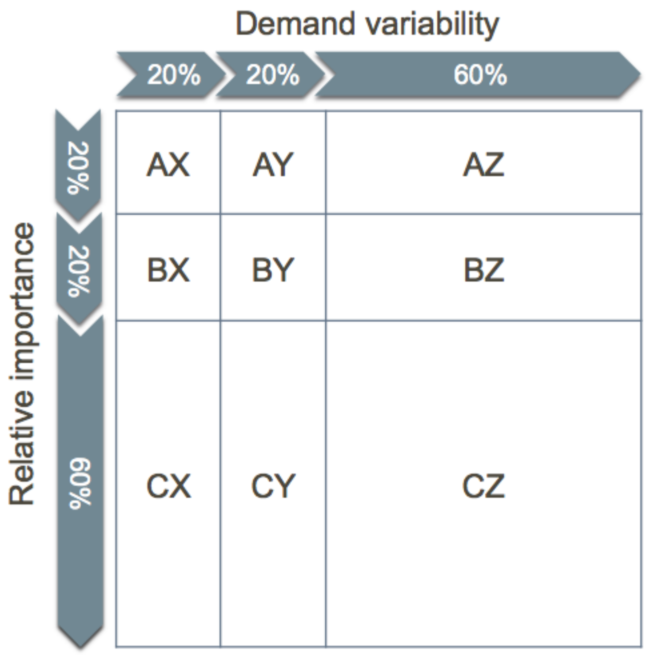

Figure 4.9: Visual representation of groups forming using 20-20-60 principle.

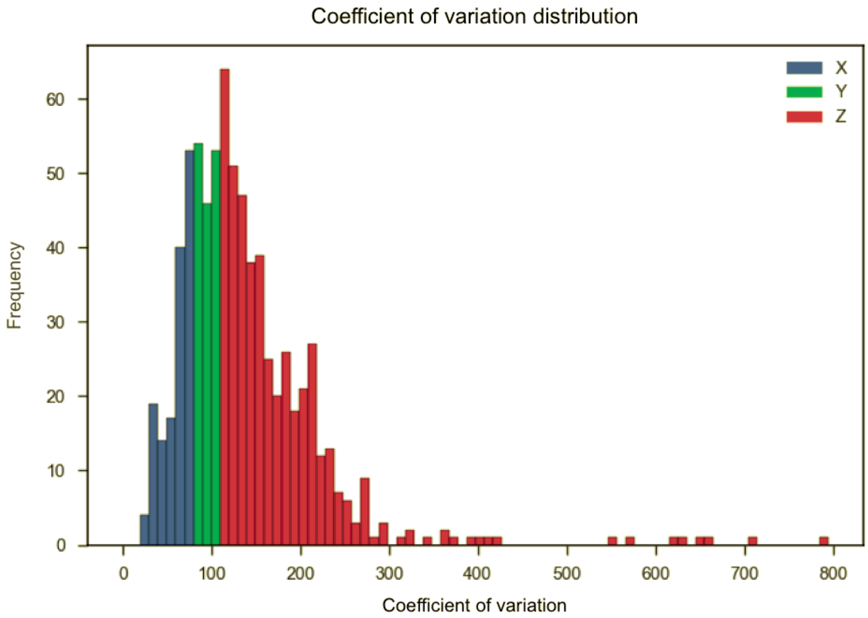

Figure 4.10: Distribution of Coefficient of Variation across parts.

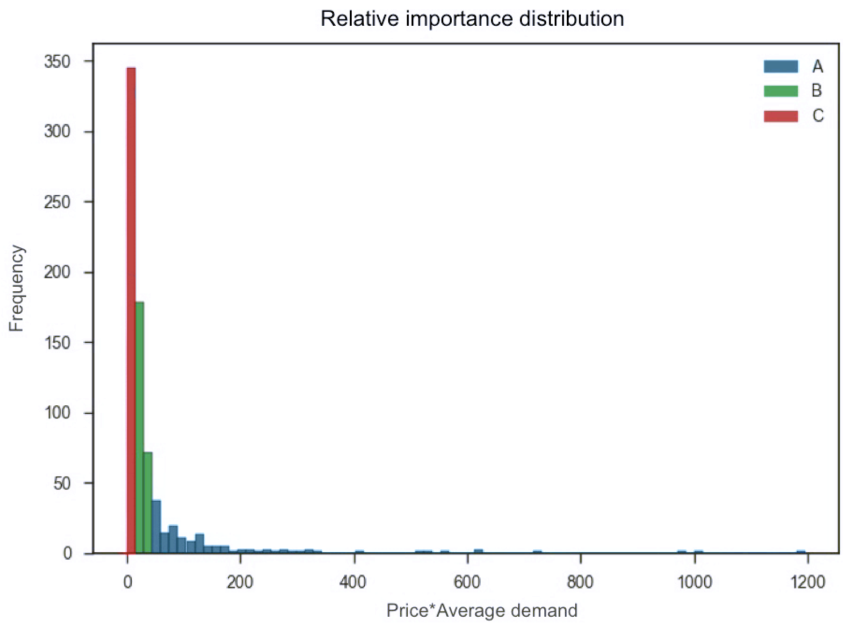

Figure 4.11: Distribution of Relative Importance across parts.

	X	Y	Z
A	82	36	31
B	35	42	73
C	32	72	346

Table 4.1: Number of parts per group.

Algorithm 3: Classification of parts

Data: Demand data of each part
Result: Dataframe for each classification group
$cov \leftarrow$ coefficient of variation
$A \leftarrow$ average daily demand
$r \leftarrow$ relative importance
$ij\_df \leftarrow$ ij group dataframe
for *each demand* **do**
    cov=(std($demand$)/mean($demand$))*100
    A=mean($demand$)
    r=A*$price$
    append these values as a row in the df
end
cdf = descending sort of df based on cov
rdf = descending sort of df based on r
$cb1$ = 0.2 * length(cdf)
$cb2$ = 0.4 * length(cdf)
$Rb1$ = 0.2 * length(rdf)
$Rb2$ = 0.4 * length(rdf)
X=cdf entries with index up to $cb1$
Y=cdf entries with index from $cb1$ to $cb2$
Z=cdf entries with index from $cb2$
A=rdf entries with index up to $rb1$
B=rdf entries with index from $rb1$ to $rb2$
C=rdf entries with index from $rb2$
for i= A,B,C **do**
    for j= X,Y,Z **do**
        ij_df = $i \cap$ j
    end
end

4.4 Statistical forecasts

Analysis of the current planning process described in the chapter 4.1 revealed flaws in the methods of prediction used by the MRP controllers. Such wrong estimates of future demand can have significant negative impact on the planning process.

Products with a great diversity of configurations available are often very difficult to forecast on the finished item level. As a solution, common practice is to forecast on the major subassembly level, where "options" in the configuration of the end product exist. The goal of this part of the thesis is to determine the best way to predict the requirements for the serial production parts. This not only means determingn the most accurate forecasting methods, but also creating the recommendations in a way that makes them easy to understand and implement. In other words, the gaol is to find an optimum between forecasting accuracy and complexity of the process.

4.4.1 Measures of error

Forecasting a variable under any level of uncertainty will always produce some errors. The goal is always to have a method that generates the smallest error possible. In order to compare performances of different forecasting methods, it is necessary to choose an adequate measure of forecast accuracy. This task can be tricky, particularly when working with data which contains a lot of zero or close to zero values. The challenge also lies in comparing across data sets that have different scales. In the course of this thesis, both of these problems were encountered. Therefore it was important to choose an appropriate measure of error accuracy. In the following lines considered measures will be discussed. The simplest explanation of forecast error is that it is the difference between the actual, real value that takes place and the value that was forecasted in a given time period. This means that the error will have a positive value when the prediction is too low, and negative when the prediction is too high.

Let A_t denote the observation at time period t and $_t$ denote a forecast of A_{t+1}.

The forecast error is then simply $e_t = A_t -_t$, which is on the same scale as the data. Accuracy measures that are based on e_i are therefore scale-dependent and cannot be used to make comparisons between series that are on different scales. [19]

The most commonly used scale-dependent measure is the mean absolute error(MAE), and root mean squared error(RMSE).

$$MAE = mean(|e_t|),$$

$$RMSE = \sqrt{mean(e_t^2)}.$$

Since this thesis aims to compare forecasting methods across many different data sets that have different scales, scale-independent measures are not adequate. Percentage errors, however, solve this problem by being scale-independent. The most commonly used one is mean absolute percentage error:

$$MAPE = mean(|p_t|),$$

$$\text{where } p_t = 100\frac{e_t}{A_t}$$

This measures has the disadvantage of being infinite or undefined if $y_i = 0$ and having an extremely skewed distribution when any y_i is close to zero. Another disadvantage (author) mentions is that it puts a heavier penalty on positive errors than on negative errors. [20]

The commonly used method is MAE/Mean ratio, often in literature abbreviated as MAD. The way it scales an error is by using the mean of the whole series, or in-sample mean. However, this ratio assumes that the mean is stable over time. As this is not true for data which show trend, seasonality, or other patterns, MAD can be unreliable. [20]

To overcome all of the disadvantages of previously mentioned measures, authors of article "Another look at the measures of forecasting accuracy" [20] propose using mean absolute scaled error (MASE).

$$MASE = mean(|q_j|),$$

$$q_j = \frac{e_j}{\frac{1}{n-1}\sum_{i=1}^{n}|A_t - A_{t-1}|}.$$

This method scales the error based on the in-sample MAE from the naive (random walk) forecast method. This makes it independent of the scale of the data. If the MASE value is less than one, it means the measured forecast has performed better than the in-sample calculated one-step naïve forecast. On the other hand, value larger than one indicates worse performance than one-step naïve forecast.

For all the mentioned reasons mean absolute scaled error (MASE) was chosen as a measure of forecast accuracy for further research.

4.4.2 Determining optimal forecasting method

Finding the best way of predicting any demand is an extensive process which requires a thorough analysis of the available data. A common practice is to start this analysis with investigating the data for the possible presence of seasonality or trend component. Discovering such patterns can often help with determining which forecasting method is most suitable and yields best results. Data visualization is usually sufficient as the first step of this process. Therefore, this analysis starts with a plot developed and optimized for visualizing any presence of monthly seasonality. Figure 4.12 shows two two examples of this plot.

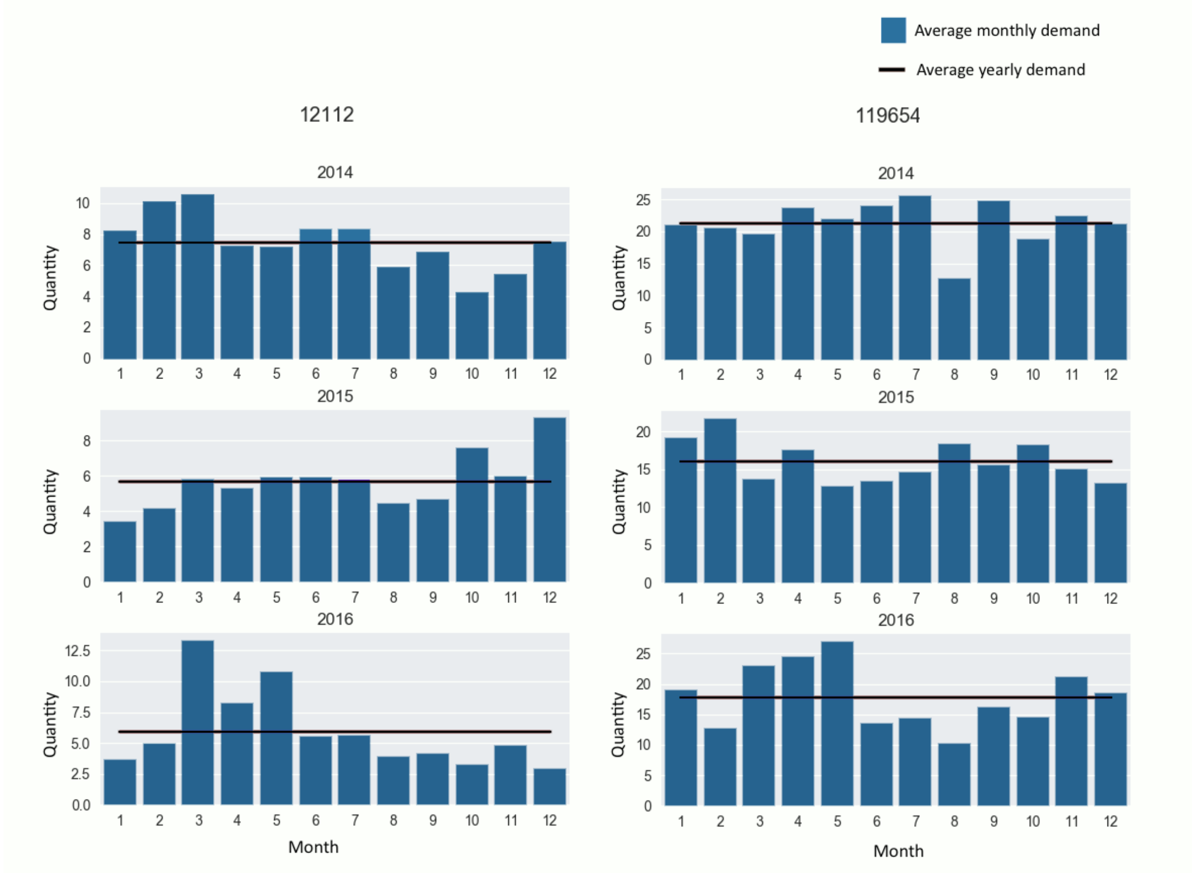

Figure 4.12: Example of plots designed for seasonality visualisation.

Reproducibility provided by the use of programming language when creating this kind of visualization allowed it to be recreated 750 times for every parts data set. Each of those plots was then visually inspected.

Blue bars in the figure 4.12 represent the average demand of a particular month while the red line represents average demand of the whole year. If monthly seasonality was present, bar heights should remain constant through years when compared with the red line. Since this is clearly not the case, it can be concluded that no seasonality exists and no further steps in that direction are required. This also means there is no reason to test forecasting methods which assume seasonality so they can be excluded from the analysis.

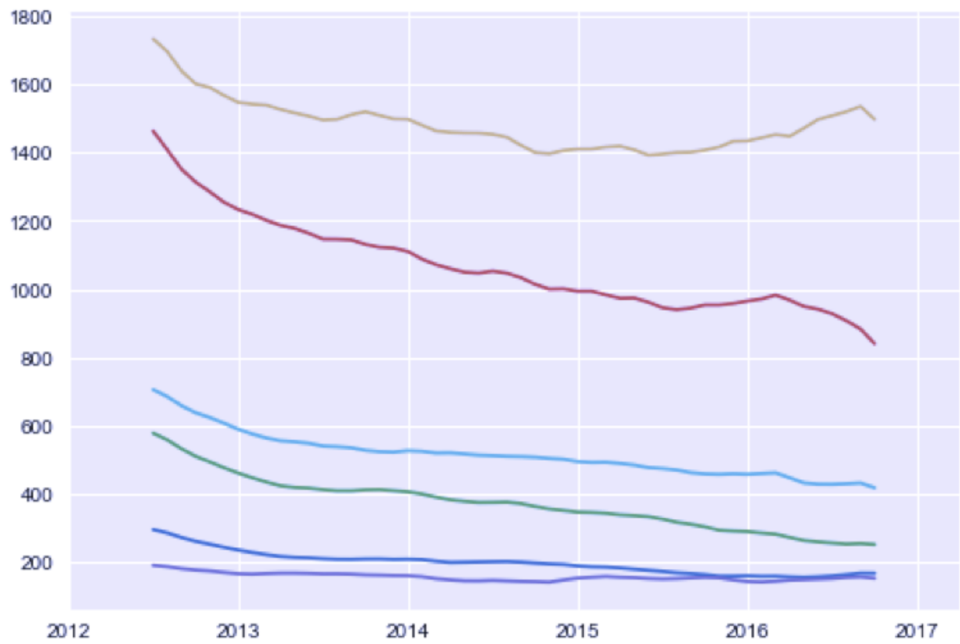

Figure 4.13: Visual representation of trend components in some parts.

Trend component is visible in some parts demand data, but only when observed over multiple years (figure 4.13). Interviews revealed that this is most probably a result of a gradual decrease in sales over the years, which doesn't have to continue in the future. However, because of occasional appearance of trend component, it was decided that trend-adjusted forecasting method should be taken into account in this analysis.

To sum up, methods which were included in the analysis are following:

- Moving average (period n ranging from 1 to 30)
- Simple exponential smoothing (α ranging from 0.1 to 0.9)
- Second order exponential smoothing(α ranging from 0.1 to 0.9 and β ranging from 0.1 to 0.9)

These methods were tested on each of the 750 parts individually. Doing it manually was almost an impossible task which would require a great effort and a lot of time. Therefore, this process was also automated using python programming language. The written algorithm works in a way that it collects the demand data of a part and then loops through all of the methods and corresponding parameters. In every loop it tests the performance of a method on the dataset. The performance is expressed in the mean absolute scaled error (described previous chapter 4.4.2). This is done for each of the 120 possible forecasting methods whose errors are then compared. At the end, the one which produces the smallest error is exported to the excel report. The process is repeated until the best forecasting method for each individual part is determined.

Clearly, using different forecasting method for each part is too complex. On the other hand, grouping parts based on the most suitable forecasting method reduces the complexity and makes it much easier to create a recommendation for the company. Therefore, the 120 possible combinations of forecasting methods and parameters were reduced to 5 and the procedure was repeated. These methods are shown in the table 4.4.2.

Next step was to calculate the performance of these substituted forecasting methods in order to quantify the potential for improvement. As previously described in the chapter 4.1 MRP controllers use many different means of predicting future demand. This makes it very challenging to determine by which degree these methods can improve the forecasting accuracy. ZSFB logic used by the SAP by default can be equated with MA_{11}. Therefore, a decision was made to use errors of this method as reference values. Table 4.4.2 shows the improvements when using different methods compared to MA_{11}. Obviously, value in the first column is very low because parameter in this method is very close to the one used by the SAP.

Method	Number of parts	Percent improvement
MA_{10}	29	0%
MA_{20}	135	1.36%
MA_{30}	352	3.01%
SES* ($\alpha = 0.1$)	125	2.38%
SOES** ($\alpha = 0.1$) ($\beta = 0.5$)	107	2.76%

* *Simple Exponential Smoothing
** **Second Order Exponential Smoothing

Table 4.2: Improvements with reduced number of forecasting methods.

Evidently, it is possible to improve the forecasting done by the SAP by choosing optimal methods and parameters. However, this can be done only by a small degree. The main reason is that the demand data of the parts shows no evident pattern. As it can be seen in the figure 4.4 it is most often normally or even randomly distributed around the average value. Nevertheless, adjusting forecasting methods and parameters in the SAP PP is straightforward and effortless process meaning that the company can implement these findings immediately as a short-term solution.

It is worth mentioning that this study also identified opportunities for more significant improvements in the area of demand forecasting. However, attaining them would require much more resources. Take as an example the so-called "frozen period." This is a period in which customer still has a chance to make changes after an order has been placed. These purchase agreements already exist but are not loaded into the system and sent to the production until the two-week frozen period expires. It was concluded that loading unconfirmed orders into the system could improve production planning by a large degree. However, interviews with stakeholders revealed that this would mean making extensive changes to the COE system, which is a time-consuming process.

Algorithm 4: Finding the best forecasting method for each data set

```
Data: Time series data
Result: Forecasting method with highest accuracy
marray ← array containg names of the forecast and corresponding MASE
  values
alpha ← smoothing constant or forecasting window
beta ← smoothing constant
for each time series do
    for each forecasting method do
        if method = SOES then
            for α = 0.1 to 0.9 do
                for β = 0.1 to 0.9 do
                    implement the forecasting method using α and β
                    calculate MASE and update marray
                end
            end
        end
        if method = SES then
            for α = 0.1 to 0.9 do
                implement the forecasting method using α
                calculate MASE and update marray
            end
        end
        if method = MA then
            for alpha = 1 to 30 do
                implement the forecasting method using alpha alpha
                calculate MASE and update marray
            end
        end
    end
    determine forecasting which produced smallest MASE within marray
end
```

4.5 Inventory management

To search for improvement possibilities in the area of inventory management it was necessary to know the inflow and outflow of goods over the course of time. This information, together with the known amount of initial inventory can be used to calculate inventory level at any point in time. This data was gathered as described in the chapter 4.2 , and used to calculate the inventory level for each time period by being included in the following formula:

$$I_t = I_t - 1 - A_t + P_t,$$

where I_t represents level of inventory, A_t the demand and P_t produced quantity at time period t. Calculated values allowed for visual representation of inventory movement. The time series plot was developed and observed in order to find improvement possibilities.

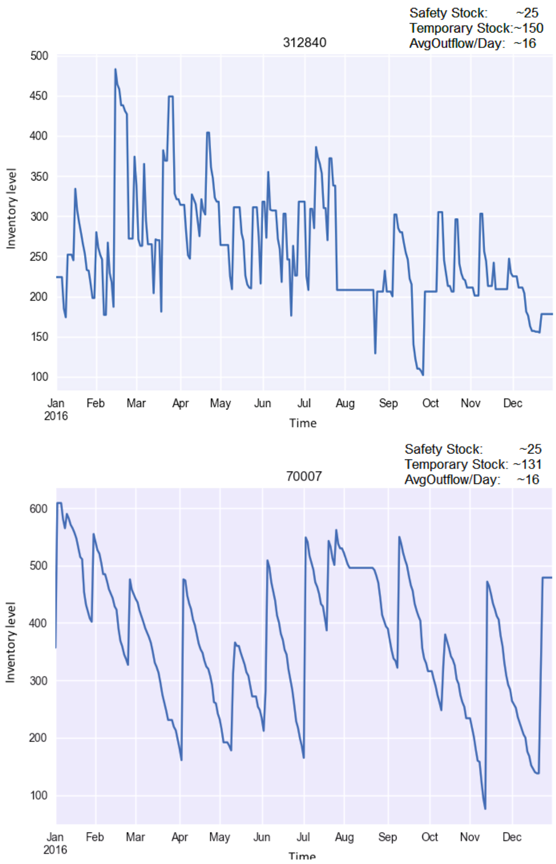

Figure 4.14: Plot representing inventory movement in one year.

When looking at the plots presented in the figure 4.14 one unusual thing can be noticed. It is clear that the safety stock level has no correlation with the variability of the demand. To prove this point, let us take a look at some examples.

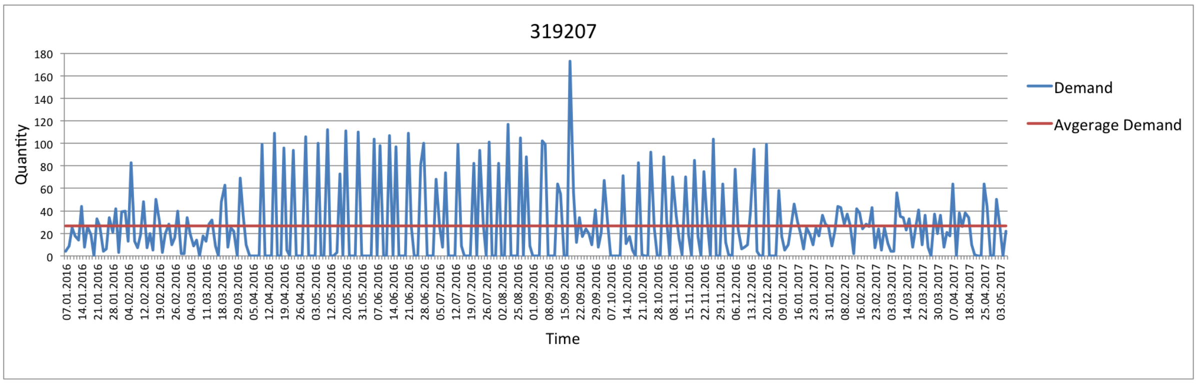

Figure 4.15: Demand for the part 319207 represented by the line chart.

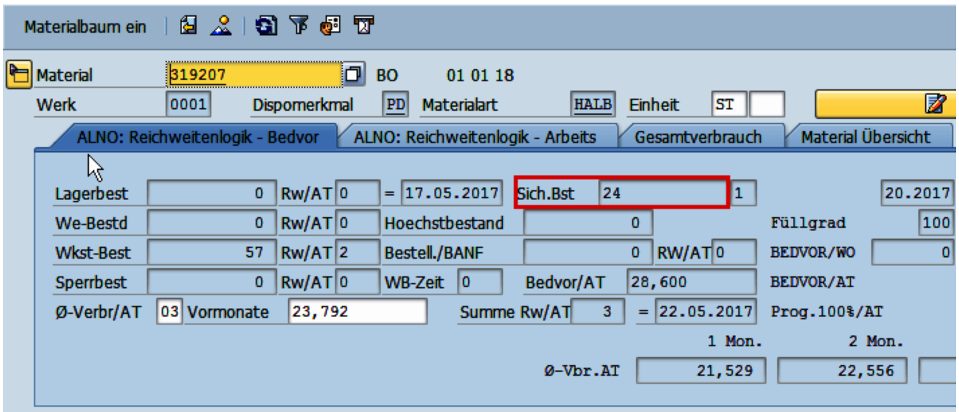

Figure 4.16: Screenshot from the SAP interface representing the part 319207 data.

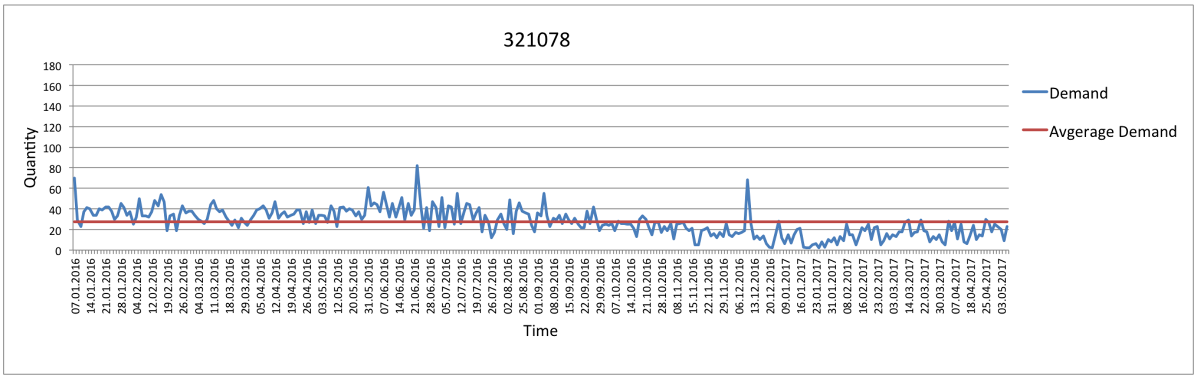

Figure 4.17: Demand for the part 321078 represented by the line chart.

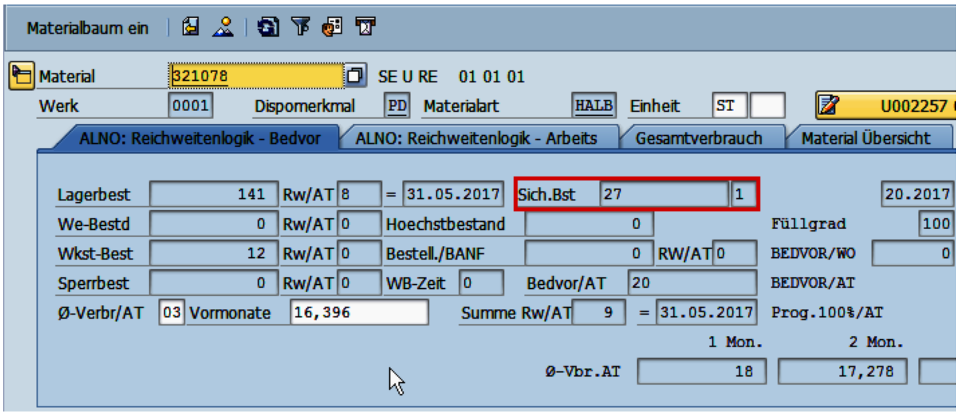

Figure 4.18: Screenshot from the SAP interface representing the part 319078 data.

Figure 4.15 shows that the part 319207 is extremely unpredictable since the demand for it varies a lot. By taking a look at the screenshot from the SAP user interface shown in the figure 4.16, it can be seen how much safety stock is reserved for it. On the other hand, demand for the part 321078 (figure 4.17) is much more stable, and its variation around the mean is less significant. However, figure 4.18 shows that almost the same amount of safety stock is reserved for this part as for the previous one. Comparing these two examples helps us prove that safety stock calculation temporary used by the company is not adequate. Hence, it was concluded that this area of inventory management leaves the largest room for improvement.

Till this day, many formulas for calculating the safety stock level are invented. However, same as with the outlier removing methods, they are intended to suit as many types of data distributions as possible, meaning they are very often suboptimal for specific cases. In addition to this, the simplified conditions assumed by these analytic models often do not apply to the real world. Solving more realistic models, on the other hand, is just too complex and time-consuming. As a way of overcoming this Ignall, Kolesar, and Walker [10] in their paper recommend using simulation.

Simulation is a process that generates a model of a real system and then tries to understand its behavior while accessing various functions through experiments. These experiments can be repeated and the results statistically processed and interpreted.

Obviously, it is unrealistic to expect from a company to run a simulation model every time they wish to determine a safety stock for a certain part. Therefore, the goal here is to use the simulation to develop a simpler analytic model and confirm it may be used with certain confidence. This is more economical as well as less time-consuming for the future analyses.

4.5.1 Simulation study

In this thesis, computer model simulation using Monte Carlo method was conducted. This method tries to replicate a random variable on the basis of knowledge about the probability distribution of that variable in the past.

In order to run a simulation, it was first necessary to establish specific inventory replenishment strategy suitable for all parts. It had to be similar to the one currently used by the company so that the model would represent a reality as accurately as possible. This meant using quantity-based (continuous review) inventory model described in the chapter 2.3.4. It was decided that some of the model parameters should be a function of the average demand A, so that they would be easy to calculate and applicable to all of the parts. Other paratmeters required by the simulation are fixed for all of the parts. Their values are represented in the following table:

Parameter	Value
Initial inventory	$10 * A$
Re-order point	$7 * A$
Order quantity	$10 * A$
Lead time	7 days
Period to simulate	1000 days
Goal service level	98%
Tolerance limit α	0,01

It is important to mention that these parameters can be easily adjusted by the company if they wish to analyse results of different scenarios and the influence of changing parameters on the system.

The main goal of the simulation was to determine adequate safety stock levels for each of the parts in order to achieve required service level. First, the algorithm 6 replicates known distribution of the demand using Monte Carlo method. It generates as many periods as necessary in order to achieve the goal of 1000 days. For example, if 400 days worth of data is available, then another 600 will be generated. After this, previously described inventory replenishment strategy is simulated for all of the periods and the number of stock-outs is determined. Knowing this number and the number of considered periods, service level can be easily calculated. Let p be the service level, i.e. the probably of not having a stock-out, then

$$p = 1 - \frac{\text{number of days with occurrence of stockout}}{\text{total number of days}}$$

A number of stock-outs depends on the level of safety stock since more safety stock means smaller probability of a stock out happening. For the first iteration, the algorithm 6 uses initially defined safety stock previously mentioned in the table 4.5.1. If the resulting service level does not equal the desired one, then a Bisection method is used to determine safety stock for the next iteration. This method is the simplest among all the numerical schemes used to find a root of a given function. To explain how it works, let us assume that the service level p can be described as a function of safety stock S:

$$p = f(S).$$

This means that determining safety stock S which will result in the target service level means finding the root of the function $f(S) - p_{tg}$, where p_{tg} represents target service level. The bisection method in this case assumes that p=p_{tg} when S is equal to some value in the given interval $[a, b]$. An algorithm computes the root of the function by repeatedly splitting this interval so that $c = (a + b)/2$. In the next iteration $[a, b]$ is replaced either by $[c, b]$ if

$$(f(a) - p_{tg}) * (f(c) - p_{tg}) > 0,$$

or with $[a, c]$ if

$$(f(a) - p_{tg}) * (f(c) - p_{tg}) < 0.$$

This process is continued until any of the following three criteria is met:

1. $|f(c_i) - p_{tg}|$ is less than predetermined tolerance limit α.
2. $|c_i - c_{i-1}|$ is less than predetermined tolerance limit α.
3. Number of iterations exceeds 100.

The algorithm 6 then moves on to the next dataset and repeats the process until service level is determined for all parts. The results are exported into the excel spreadsheet.The complete report contains following properties for each part:

- **Material number**
- **Skewness of the distribution** represented by the medcouple. This a robust statistic used to measure the skewness of a univariate distribution. Brys, Hubert, and Struyf (reference) define it as a scaled median difference of the left and right half of a distribution.
- **Group** to which the part belongs, determined by the classification process (reference).
- **Coefficient of variation**.
- **Average daily demand**.
- **Reached service level** represents the closest possible service level to the target one. If one of the previously mentioned conditions 2 or 3 is true before 1, this means it is not possible to reach the target service level p_{tg} using given parameters.
- **Safety stock** determined by the algorithm.
- **Normalized safety stock** required to compare the results across all parts. This means that safety stock values hat to be adjusted to a notionally common scale. The simple way of doing this is using formula

$$S_n = \frac{S}{A},$$

where S_n is normalized safety stock.

Algorithm 5: Simulation function

Data: Demand data and inventory policy parameters
Result: Service level for the given inputs
re_point ← reorder point
$df[x][y]$ ← dataframe entry in column x, row y

% columns description
demand ← real and simulated demand column
BI ← beginning inventory
EI ← ending inventory
UR ← units recieved
DD ← index of the day when order is recieved
LS ← lost sales
OQ ← order quantity
WFO ← waiting for order - True or False

for *i in range(0,lenth(df)* **do**
    df[WFO][i]=False
end
df['BI'][0]= *beginning_inventory*
re_point=in_re_point + safery_stock
% algorithm continues on the next page

```
re_point=in_re_point+ss
for i in range(0,lenth(df)) do
    if i > 0 then
        df [BI][i]=df[EI][i-1]
    end
    if i in df[DD] then
        df[UR][i]=df[OQ][i-lead]
    else
        df[UR][i]=0
    end
    df[EI][i]= max(df[BI][i]+ df[UR][i]-df[demand][i] ,0)
    if df[demand][i]>df[BI][i]+df[UR][i] then
        df[LS][i]=abs(df[BI][i]+df[UR][i]-df[demand][i])
    else
        df[LS][i]=0
    end
    df[EI][i]= max(mydf[BI][i]+ mydf[UR][i]-mydf[demand][i],0)
    if df[demand][i]>df[BI][i]+df[UR][i] then
        df.loc[LS][i]=abs(mydf[BI][i]+mydf[UR][i]-mydf[demand][i])
    else
        df['LS'][i]=0
    end
    if df[BI][i]<=re_point and df[WFO][i]==False then
        mydf[OQ][i]=up_point
        if i<length(df)-8 then
            for x in range(i+1,i+7) do
                df[WFO][x]=True
            end
        else
            for x in range(i+1,len(mydf)) do
                mydf[WFO][x]=True
            end
        end
        if df[OQ][i]<0 then
            df[OQ][i]=0
        end
        df[DD][i]=i+lead
    else
        df[OQ][i]=0
        df[DD][i]=0
    end
end
service_level=1-(length(df([LS != 0))/length(df))
```

4.5.2 Correlation analysis

Clearly, it is unrealistic to expect from a company to run a simulation model every time they wish to determine a safety stock for a particular part. An obvious thing to do was to use simulation study results as means of developing a simpler analytic model. This model is supposed to predict the adequate safety stock level based on some other known property. Moreover, analytic formula is more economical and less time-consuming for the future analyses. Developing it meant finding a correlation between numerically determined safety stock and one or more of the easily calculated independent variables:

- Average daily demand
- Distribution skewness
- Coefficient of variation

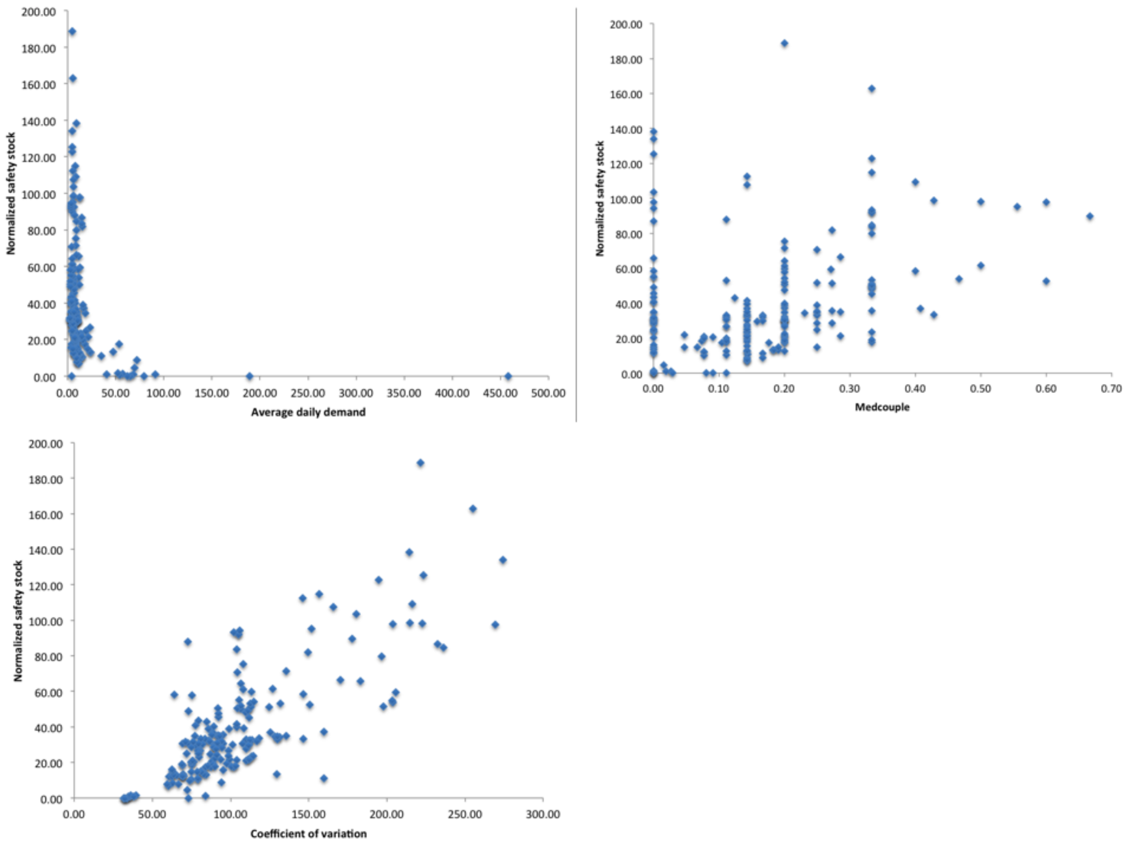

Figure 4.19: Scatter diagram representing relationships between different measures and S_n.

As mentioned before in the chapter 2 creating a scatterplot should be a preliminary step preceding any correlation analysis of the two variables. By looking at the figure 4.19 it can be assumed that only independent variable which shows shows signs of linear correlation with normalized safety stock is the coefficient of variance.

However, in order to prove it, a correlation should be quantified. Microsoft Excel provides a simple way of calculating correlation coefficient by using PEARSON function. This function simply returns the Pearson correlation coefficient between two arrays of data.

Measurement	Pearson correlation coefficient
Average demand	-0.25
Medcouple	0.35
Coefficient of variation	0.79

Table 4.3: Results of Microsoft Excel correlation analysis.

The results presented in the table 4.3 confirm that the correlation between normalized safety stock and coefficient of variance is the only one worthy of further analysis.

4.5.3 Regression analysis

Scatterplot presented in the figure 4.19 indicates that the correlation may vary across the C_v range. Because of this, it made sense to first investigate if the strength of the correlation is different across different classes of products. In the chapter 4.3 parts were split up into categories X, Y and Z based on the demand coefficient of variance. If these different categories show different correlation with normalized safety stock, this classification should be used again in order to establish an individual rule for each category.

A decision was made that a line of best fit should be determined independently for X, Y and Z parts. To do this, Microsoft Excel provides a simple way of utilizing the method of least squares previously described in the chapter 2. Results of this analysis are presented by the table 4.4.

Group	C_v Range	Fitted equation	R^2
All together	0 - 274.18	$y = 0.547x - 18.078$	0.619
X	$0 - 79.65$	$y = 0.614x - 21.427$	0.297
Y	$79.65 - 108.95$	$y = 1.349x - 91.167$	0.324
Z	$108.95 - 274.18$	$y = 0.619x - 32.689$	0.564

Table 4.4: Results of Microsoft Excel regression analysis.

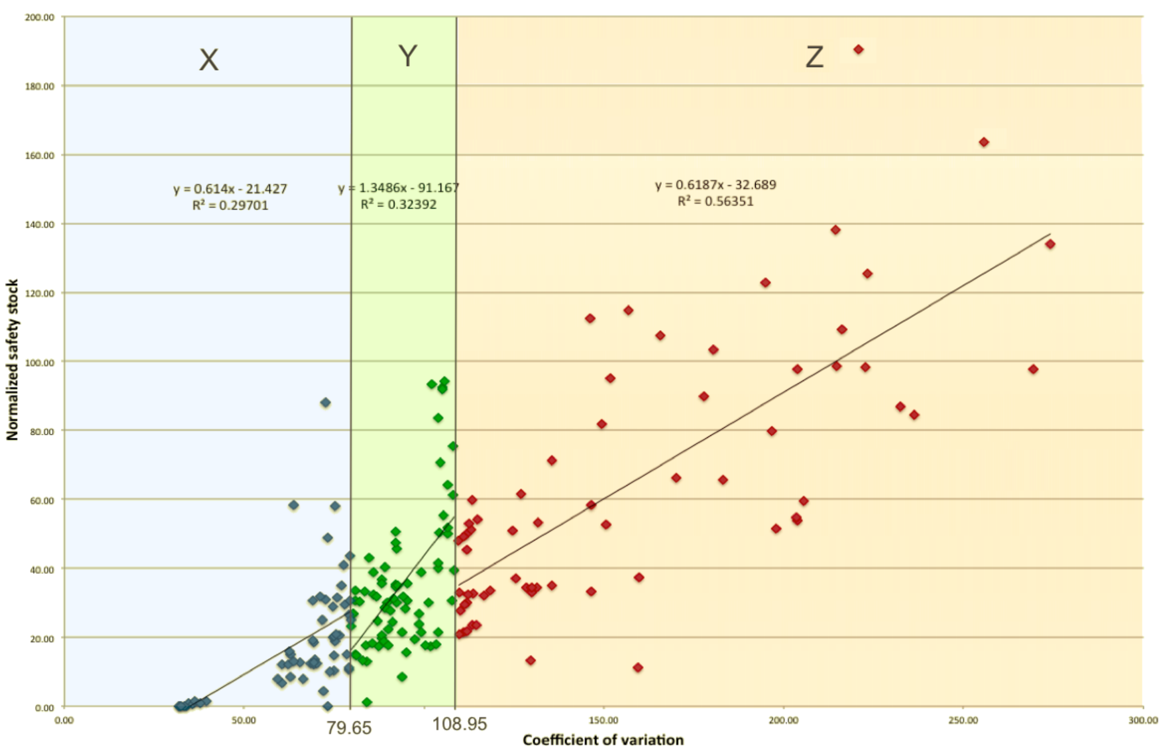

Figure 4.20: Scatterplot together with individual lines of best fit for each category.

As it can be seen in the table 4.4, R^2 equals 0.564 for class Z, which is a satisfactory fit. It means that 56 % of the variation in normalized safety stock is explained by the independent variable coefficient of variation. The closer to 1, the better the regression line fits the data. Correlations in classes X and Y are not as significant, so fitted equations for these two classes may not be an optimal means of determining adequate safety stock. Instead, using average normalized safety stock of a class as an approximate value may be more convenient in this case.

Group	C_v Range	Method of determination	Normalized safety stock
X	$0 - 79.65$	Average of the category	18.6
Y	$79.65 - 108.95$	Average of the category	35.2
Z	$108.95 - 274.1$	Analytical formula	$0.62 * C_v - 32.69$

Table 4.5: Recommendations for determining normalized safety stock.

Determining safety stock amount for a particular part is then as simple as multiplying calculated normalized safety stock with the average demand.

Evidently, the combination of simulation and regression analysis can be a powerful means for the developing analytical formula. Combining the findings obtained by these methods with the information on how much resources are designated for production planning allowed the creation of the optimal safety stock determination procedure.

4.6 Case study conclusion

Previous segments of the case study research chapter described the process of data analysis which yielded some significant results. This section will briefly summarize these results and combine them with a goal of creating overall recommendations for the company.

4.6.1 Results

When improving production planning, inventory management and the allocation of the resources reserved for these tasks, the first logical step was to classify parts based on some of their properties. As described in the section 4.3 of this chapter, relative importance of the part and variability of its demand were chosen as the most significant. Parts were then split into nine groups based on these two measures. This classification will be used as a guideline for applying different planning and inventory management strategies on different groups of parts.

Analysis of the current prediction making process explained how in some cases methods used by MRP controllers give a wrong picture of reality and consequently lead to incorrect estimations of future requirements. Due to the fact that the company's standard forecasting process doesn't take historical demand into account, a possibility of using statistical forecasting methods was investigated. This analysis confirmed that improvements can be made by choosing different methods and different time periods (i.e. lags). Compared to the SAP forecasting, methods with optimized parameters showed an average of 2.4 % error decrease per part, whereas reducing the number of possible forecasting methods to 5 resulted in an average of 2.4 % decrease. By comparing tables 4.8 and 4.7 it can be concluded that the accuracy of the fitted forecasts is in almost all cases proportional to the coefficient of variance.

Analysis of the company's inventory management, specifically stock policy, also showed significant room for improvement. Monte Carlo simulation of (s,q) inventory strategy was conducted on a sample of 200 parts. The results were then used to determine how much safety stock is required to achieve a 98% service level of each part. Finally, linear regression allowed determining equations which best describe a correlation between an independent variable and the numerically calculated safety stock. The coefficient of determination R^2 was used to quantify the goodness-of-fit of these equations. It showed that the coefficient of variation describes aproximately 30

4.6.2 Recommendations

The results can be of great value to the company if applied strategically. A convenient way of structuring production planning resource allocation is by using the classification. Following lines will briefly explain which approach is suitable for which group and why. Tables 4.6, 4.7 and 4.8 show the parameters important for understanding the reasons behind assigning specific strategy to a specific group.

	X	Y	Z
A	43.74	74.08	162.09
B	48.84	81.40	134.89
C	58.80	81.17	173.04

Table 4.6: Average coefficient of variance of each group expressed in percentage.

	X	Y	Z
A	38.59	156.50	106.07
B	53.46	35.94	45.06
C	18.43	16.29	11.01

Table 4.7: Average relative importance (i.e. daily revenue) of each group expressed in euros.

	X	Y	Z
A	48.84	107.51	146.98
B	52.38	95.08	130.35
C	59.84	82.47	79.12

Table 4.8: Average MAPE of five fitted statistical forecasting methods for each group, expressed in percentage.

AX, BX, CX

Demand for parts which belong here is stable and thus easy to predict. In the table 4.8 it can be seen that MAPE produced by statistical forecasting methods is relatively small for these groups. Hence, fitted methods are sufficient to make reliable predictions of the future demand. Eventual fluctuations can easily be covered by utilizing the safety stock whose calculation should be based on the average normalized safety stock as explaind in the section 4.5. The process of production planning can safely be automized and requires very little control.

AY,BY

The attention should be shifted to the parts belonging to these groups. They are important but have less stable demand which makes them unsuitable for solely automized planning. The statistical forecasting methods can't entirely be entrusted with making predictions for their future demand because MAPE in these cases is noticably larger. Instead, the recommendation is to plan the production of these parts by strategically combining the experience of MRP controllers and statistical forecast. Their planning, therefore, requires more resources and attention. Safety stock calculation should also be based on the average normalized safety stock of the group.

AZ

These parts are extremely unpredictable while being very important for the company and expensive to hold in inventory. A recommended solution is to make these parts only to order (MTO), which would reduce the risks of stock-outs or excessive inventory. Only 31 part belongs to this category and planning them separately from the other parts wouldn't require too much effort from the MRP controllers. However, if safety stock is still necessary, it should be calculated using fitted equation as described in the section 4.5.

BZ

BZ parts are also very unpredictable and still relatively important. However, compared to AZ parts, holding them on stock is less expensive. Their demand can be forecasted using selected statistical forecasting methods. However, to ensure service level of approximately 98%, the calculation of their safety stock should be done using developed analytical formula.

CY, CZ

Holding parts belonging to the CY and CZ groups on stock is relatively inexpensive. They are either cheap or required in small quantities which gives them low relative importance value. Therefore the recommendation is to hold enough of them in the inventory so that they require less attention from the MRP controllers.

5 Conclusion

Over the course of time, growing competition in the kitchen manufacturing industry forced the studied company to continuously provide its customers with more customisation possibilities. This led to the more expensive and complex production planning as well inventory management process. As a result, the need for higher effectiveness and efficiency arose.

This thesis resulted from months of active participation in this company, during which time many different opportunities for improvement were identified. It is worth mentioning, however, that only the most important ones were further investigated and included in this paper.

Discovering a large amount of historical data stored by the company's ERP system revealed it's great potential to advance the production planning and inventory management process. Further extensive work which mainly involved processing and analyzing these records of historical demand resulted in a valuable knowledge about manufactured parts. The research, therefore, bases itself mainly on data analysis and mathematical computations which were conducted by utilizing the advantages of computer algorithms.

That being said, this case study and its findings serve as a great addition to the multiple lines of evidence which support the incentive of using data analysis in today's manufacturing. They prove that by strategically using this data-driven knowledge, companies can successfuly optimize their production planning and inventory management process to the otherwise unreachable extent.

In addition to this, this thesis strives to further the understanding of the whole procedure which led to the results. Applying different methods and tools was done in a structured way and can easily be replicated in other companies and organizations.

Bibliography

[1] Alno AG. *Annual Report*. 2015. URL: http://www.alno.de (cit. on p. 22).

[2] Howard S. Becker. *Sociological Work: Method and Substance.* Adline, Chicago, 1970 (cit. on p. 3).

[3] Stephen N. Chapman. *The Fundamentals of Production Planning and Control*. Pearson Prentice Hall, 2006 (cit. on pp. 7–11, 15, 16, 18).

[4] John H. Cochrane. *Time Series for Macroeconomics and Finance*. Graduate School of Business, University of Chicago, Chicago, 1997 (cit. on p. 9).

[5] Peter J. Diggle. *Time Series: A Biostatistical Introduction, Oxford science publications*. Clarendon Press, 1990 (cit. on pp. 9, 11).

[6] Raya Fidel. "The Case Study Method: A Case Study." In: (2013) (cit. on p. 3).

[7] Michael Friendly and Daniel Denis. "The Early Origins And Development Of the Scatterplot." In: (2005) (cit. on p. 18).

[8] D.M. Hawkins. *Identification of Outliers*. Chapman and Hall, London – New York, 1980 (cit. on p. 36).

[9] Honeywell. *Manufacturers Rely On Big Data Analytics To Navigate Tough Business Conditions*. URL: https://www.honeywell.com/newsroom/pressreleases/2016/09/manufacturers-rely-on-big-data-analytics-to-navigate-tough-business-conditions (cit. on p. 2).

[10] Edward J. Ignall, Peter Kolesar, and Warren E. Walker. "Using Simulation to Develop and Validate Analytic Models: Some Case Studies." In: (1977) (cit. on p. 57).

[11] Deloitte Touche Tohmatsu Limited. *Analytics Advantage Survey*. 2013. URL: https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Deloitte-Analytics/dttl-analytics-analytics-advantage-report-061913.pdf (cit. on p. 2).

[12] M.Hubert and E. Vandervieren. "An Adjusted Boxplot for Skewed Distributions." In: (2008) (cit. on p. 37).

[13] Encyclopedia of Mathematics M.S. Nikulin (originator). *Three-sigma rule*. URL: http://www.encyclopediaofmath.org/index.php?title=Three-sigma_rule&oldid=17366 (cit. on p. 37).

[14] NIST/SEMATECH. *e-Handbook of Statistical Methods*. 2003. URL: http://www.itl.nist.gov/div898/handbook/ (cit. on p. 36).

[15] Roxy Peck, Chris Olsen, and Jay L. Devore. *Introduction to Statistics and Data Analysis*. fifth edition. Cengage Learing, 2016 (cit. on pp. 19, 20).

[16] B. Joseph Pine. *Mass Customization: The New Frontier in Business Competition*. Harvard Business School Pres, 1990 (cit. on p. 26).

[17] Gartner research. *Dark Data*. URL: https://www.gartner.com/it-glossary/dark-data (cit. on p. 1).

[18] Gartner research. "Predicts 2016: Opportunities Abound for the Factory to Reach Its Potential." In: (2015) (cit. on p. 1).

[19] G. Athanasopoulos R.J.Hyndman. *Forecasting: principles and practice*. OTexts, Melbourne, Australia., 1990 (cit. on pp. 14, 47).

[20] Anne B. Koehler Rob J. Hyndman. "Another Look at Measures of Forecast Accuracy." In: (Nov. 2005) (cit. on p. 48).

[21] *SAP Documentation*. 2009. URL: http://fmis.minfin.bg/help/EN/3f/ceecc56f6511d3be87080009b31c68/frameset.htm (cit. on pp. 25, 26).

[22] William J. Stevenson. *Operations Management*. eleventh edition. McGraw-Hill Irwin, 2012 (cit. on pp. 8, 12, 13, 15, 18).

[23] Dileep R. Sule. *Production Planning and Industrial Scheduling*. second edition. CRC Press, Taylor Francis Group, 2008 (cit. on pp. 1, 7, 9, 14).

[24] Robert K Yin. *Case study research: design and methods*. fourth edition. Sage Publications, Thousand Oaks, 2009 (cit. on p. 3).