Improving Forecasting for the Aftermarket through Big Data

A Case Study at Volvo Group

Master of Science Thesis
in the Supply Chain Management master's Programme

RAHAND NAWZAR
SAMI KARLSSON SHEIK

Department of Technology Management and Economics
Division of Service Management and Logistics
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2016
Report No. E2016:072

REPORT NO. E2016:072

Improving Forecasting for the Aftermarket through Big Data

A Case Study at Volvo Group

RAHAND NAWZAR

SAMI KARLSSON SHEIK

Tutor, Chalmers: Dr. Henrik Ringsberg
Tutor, Volvo Group: Marcus Bohman

Department of Technology Management and Economics
Division of Service Management and Logistics
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden 2016

Improving Forecasting for the Aftermarket through Big Data
A Case Study at Volvo Group
RAHAND NAWZAR
SAMI KARLSSON SHEIK

Report No. E2016:072

Department of Technology Management and Economics
Division of Service Management and Logistics
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
Telephone: + 46 (0)31-772 1000

Chalmers Reproservice
Göteborg, Sweden 2016

Improving Forecasting for the Aftermarket through Big Data
A Case Study at Volvo Group
Rahand Nawzar and Sami Karlsson Sheik
Department of Technology Management and Economics
Division of Service Management and Logistics
Chalmers University of Technology

ABSTRACT

The aftermarket brings profitable advantages to manufacturing companies by providing value adding services for the costumers, where accurate forecasting is essential in order to achieve a smooth material flow of spare parts. Concurrent forecasting methods are mainly based on historical demand with mathematical methods that can trace back to the 40s. Today's business setting where companies handle large amounts of data, also known as big data, provides new and innovative improvement possibilities. Naturally, spare parts demand is difficult to monitory since occurrence of failure is unpredictable. Forecasting based on big data might be a way to achieve high uptime for the costumer by having the spare parts at the right place in the right time, addressing the challenge of having high availability at a low cost. The purpose of this thesis where to investigate opportunities for improving forecast accuracy of spare parts in the aftermarket of automotive companies, by exploiting big data created downstream the supply chain.

A case study was conducted at Volvo Group, which is a market leading automotive manufacturer. Two distinct research questions were identified in order to fulfill the purpose of the thesis; 1) which data created in an aftermarket supply chain has potential to increase the accuracy in predicting demand of spare parts, and 2) how can the identified data support planning processes of automotive companies for predicting demand of spare parts in the aftermarket. Volvo Groups aftermarket supply chain was scrutinized in order to provide insight for potential opportunities of big data utilization. In combination with a theoretical framework which provides academic insight into concurrent research the defined research questions were answered.

The result of this thesis is a framework which describes two dimensions for succeeding in implementing big data in the planning process of spare parts. The dimensions are presented through matrices which gives an illustrative view on how big data is utilized in the planning process of spare parts. The first dimension describes the level of sophistication to translate big data into a demand. The second dimension describes the level of integration needed towards the department responsible for forecasting. In conclusion, this framework has the potential to guide not only the automotive industry but also other industries that has access to big data which correlates to the probability that a spare part fails.

Keywords: forecasting, big data, spare parts, aftermarket, supply chain visibility, regression analysis and life data analysis

ACKNOWLEDGEMENT

Firstly, we would like to thank our supervisors, Dr. Henrik Ringsberg from the institution of Technology Management and Economics at Chalmers University and Marcus Bohman from Volvo Group. Your passionate nature for research and academia has been an inspiration. We are truly grateful and thankful for the guidance and extensive support we have received throughout the thesis.

Additionally, we express appreciation for the enthusiastic support provided by employees at Volvo Group. Your participation in interviews and meetings has been a critical part of our success. We thank you for the immense knowledge which you have shared with us.

Finally, we thank our friends and family for the continuous support and care we have received throughout our years at Chalmers University. All of which has lead us to this master thesis and thus it is dedicated to you all.

Gothenburg 2016,

Rahand Nawzar & Sami Karlsson Sheik

TERMINOLOGY

This section presents definitions of terminology used throughout the thesis, references can be viewed in Appendix C.

Subject	Definition
Supply chain management	The management of material, information and capital in a supply chain, which aims at providing the end-customer with maximum satisfaction at the lowest cost.
Logistics	The management of material or other resources which is in movement or at rest in certain locations.
Aftermarket	An after-sales market which aims to provide customers with goods and services that are necessary for maintaining the functionality of a previously sold product.
Automotive	Revolves around the design, development, manufacturing, promotion and retailing of motor driven vehicles.
Spare parts	An identical replacement of a component, used when the original component needs to be replaced.
Inventory	Materials and goods which are held by a company to support different business processes, such as production.
Big Data	Very large and complex datasets which can't be handled by traditional data processing methods.
Connectivity	Nodes in a network which are connected to each other and can communicate through the network.
Predictive Analytics	Using mathematical models for analyzing large datasets in order to make predictions of the future.
Forecasting	Tool for managing the uncertainty in future demand within a certain time frame.
Forecast accuracy	A process in which the quality of the conducted forecasts is assessed.
Supply chain visibility	Explained as the information collected and shared between actors or nodes in a supply chain.
Relationships	Connection between different factors, both tangible and intangible.
Linear regression	Mathematical technique of creating a linear function from data points which can be used to estimate future values.
Applied life data analysis	The analysis and prediction of products life lengths.

ABBREVIATIONS

IT	Information technology
VG	Volvo Group
MM	Material Management
S&OP	Sales and Operation Planning
GTO	Volvo Group Truck Operations
GTT	Volvo Group Truck Technology
POS	Point of Sales
KPI	Key Performance Indicators
SKU	Stock Keeping Units
MAPE	Mean Absolute Percentage Error
LS	Logistics Services
CDC	Central Distribution Centers
SDC	Support Distribution Centers
RDC	Regional Distribution Centers
DIP	Demand and Inventory Planning
CMP	Continental Material Planning
Refill	Refill Material Management
DIM	Dealer Inventory Management
LPA	Logistics Partnership Agreement
SS	Safety Stock
EOQ	Economic Order Quantity
ROP	Reorder Point
OEM	Original Equipment Manufacturers
VT	Volvo Trucks
PDM	Product Data Management Systems
BOM	Bill of Material

TABLE OF CONTENT

TABLE INDEX

FIGURE INDEX

EQUATION INDEX

1. INTRODUCTION

Initially, this chapter describes a background for the master thesis. Where a broad description of current opportunities for big data in forecasting and challenges in the aftermarket are presented. The issues with said areas are then explained and the purpose and research questions of the thesis is stated and discussed. Finally, the thesis structure is presented with brief descriptions for each chapter.

1.1 BACKGROUND

We are in the midst of a revolution, our world is changing in a rapid pace and the driving force is called big data (Lohr, 2012). As one of many concurrent buzzwords the big data concept is not only a trend but also a concept which offers significant opportunities to change modern business models and day-to-day decisions (Waller and Fawcett, 2013).

The immense sources of data which is available through our increasingly interconnected world will have a huge impact on societal institutions, be it businesses or governments (Lohr, 2012; World Economic Forum, 2012). For businesses it has always been a common view that data is a driver for increasing profitability and facilitating decision making (Waller and Fawcett, 2013). Published research also shows that one-third of all top-field companies that use big data for managerial decisions, are six percent more profitable than their competitors (Mcafee et al., 2012; Lohr, 2012). An example of how companies can simplify and support managerial decisions is by sharing data throughout the entire supply chain (McIntire, 2014). In addition, exploitation and utilization of data increase a company's competitiveness through a more visible supply chain (McIntire, 2014).

A visible supply chain may help predict upcoming events and support better planning decisions, which directly affect the effectiveness of a company's supply chain (McIntire, 2014). However, companies that aim to increase their competitiveness in the future must invest in the development of tools which exploits big data, today (Hassani and Silvia, 2015). One of these tools, as mentioned by Hassani and Silvia (2015), is big data's potential for improving forecasts. Essentially forecasting is used to manage uncertainty in demand by predicting the future (Waller and Fawcett, 2013). As emphasized by Ton (2013) forecasting and big data goes hand in hand and can be improved by exploiting the increasingly available data within the focal company.

Uncertainty in demand for spare part derives from huge number of stock keeping units (SKU) (Cohen et al., 2006; Jouni et al., 2011). This can be mitigated through high inventory levels or responsive transportation solutions. Such approaches for dealing with uncertain demand are associated with high costs (Jouni et al., 2011). Naturally, improving forecasts should have a significant impact in securing spare parts at the right place, in the right time and for the right cost. In effect it is possible to decreasing inventory costs without impacting the availability negatively. The availability of spare part is increasingly important for the automotive industry because it is moving towards providing solutions and services to its end-customer, rather than just selling a product (Cohen et al., 2006). This point of view brings multiple benefits such as increased profitability and customer loyalty. It has been found that aftermarket services can triple the turnover for new products (Bundschuh and Dezvane, 2003; Gaiardelli et al., 2007; Saccani et al., 2007). Thus it is of true significance to minimize the occurrence of stock-outs which could decrease the uptime of a vehicle; if a stock-out occurs this could result in a direct loss of sale for the focal company (Kennedy et al., 2002; Wang and Syntetos, 2011). Additionally, improving forecast accuracy for the aftermarket has the possibility of decreasing costs related to keeping parts in inventory without impacting the availability negatively (Romeijnders et al., 2012).

Automotive companies are collecting data from end-customers at an increasing rate and it is vital to investigate how to utilize such data (Frowein et al., 2014). From a material management perspective it can help facilitate and manage the customer's demand for uptime and improve material flow. Thus providing greater customer value to compete more fiercely in the aftermarket. The automotive industry is facing an increase in demand of higher service levels from end-customers (Robinson, 2014). The increasing amounts of data which is created, through for example connected vehicles, give a potential to capture new business opportunities and translate it into customer value (Frowein et al., 2014). By exploiting such data and incorporating it into current forecasting methods there is great potential for cost saving for the focal company and value creation for the end-customer.

When considering the importance of accurately predicting demand for spare parts and the fact that forecasting is a subject which has been extensively researched since the 1970s. There is a lack of contemporary research in the described areas, which consider the increasing amount of big data generated from a more interconnected world. This thesis gives insight as to the usage of big data in forecasting for the aftermarket. It shows, through a case study, how big data can facilitate the planning processes of spare parts and thus improve forecast accuracy.

1.3 PURPOSE AND RESEARCH QUESTIONS

The purpose of this master thesis is to investigate the usage of big data created downstream the supply chain at automotive companies for improving forecast accuracy of spare parts in the aftermarket.

Firstly, to meet the stated purpose a general view of the potential data parameters which can be used to increase forecast accuracy needs to be investigated. Such knowledge is critical in order to successfully understand and analyze how different datasets can be combined to improve forecast accuracy. Secondly, by scrutinizing current demand planning processes of spare parts at automotive manufacturers it is possible to identify feasible ways to improve forecasting through big data. This is increasingly important because not only must potential data parameters be identified but also integrated, thus such an investigation provides insight into feasible ways for data integration.

The following research questions have been formulated in order to achieve the purpose of this master thesis:

RQ1: Which data created downstream in an automotive aftermarket supply chain has potential to increase the accuracy in predicting demand of spare parts?

RQ2: How can the identified data support the planning processes of automotive companies for predicting demand of spare parts in the aftermarket?

1.4 SCOPE AND LIMITATIONS

Because of the complex nature of big data, in terms of technological sophistication, this study excludes monetary effects of integrating big data in planning processes for spare part. The scope regarding integration lies in the complexity of making it feasible. Furthermore, big data in this study is limited to any data which is created downstream the supply chain of an automotive company; this includes any data collected or received from end-customers. This study does not consider how changes in forecast accuracy of spare parts affect the material flow.

1.5 THESIS STRUCTURE

The thesis is structured as follows:

Methodology	Explains how the thesis work was conducted; how the theoretical framework was built and how empirical findings were extracted. It presents the chosen research strategy of the thesis, how empirical evidence was gathers and the studies validity and reliability.
Theoretical framework	This chapter describes the studied literature which is to be used as basis for the analysis. Areas which are relevant in order to fulfill the purpose are presented. Such as: the aftermarket, spare parts, forecasting, big data and supply chain visibility. Finally, a conclusion of the theoretical study is presented.
Empirical findings	In this section all of the empirical findings will be presented. General information about Volvo Group as a global automotive manufacturer, the construction of their aftermarket supply chain, how planning processes are executed and which data parameters are created downstream the supply chain.
Results and analysis	This section initially give a description of big data's potential in improving forecasting. Additionally, a two dimensional framework is presented which describes the levels of integration in the planning process of spare parts and levels of sophistication for translating big data into a demand.
Conclusion	This chapter highlights significant findings from this study. A summarization of the developed framework is presented, in addition to elaborating regarding the two dimensions.
Discussion	This section elaborates on strengths and weaknesses in the study. It proposes areas of interest for future research and describe contributions, both theoretical and practical.

2. METHODOLOGY

This chapter presents how the thesis work has been executed; it explains the chosen research strategy which was used for this thesis. Furthermore, it describes which methods were used for gathering empirical evidence followed by a discussion of the study's validity and reliability.

2.1 RESEARCH STRATEGY

When choosing research strategy there are two alignments, either conducting a quantitative or qualitative study (Bryman and Bell, 2015). The main difference being that the former is based on measuring or collecting data subject to an analysis and the latter is based on a subjective analysis with measurements that can't be quantified (Bryman and Bell, 2015). However, there is also the possibility of conducting a mixed methods approach which combines both a qualitative and quantitative methods (Dubois and Gadde, 2002). A common mixed research strategy is called systematic combining, which combines a review of theory and empirical findings in an iterative process (Dubois and Gadde, 2002). According to Dubois and Gadde (2002) systematic combining will expand the understanding of both theoretical and empirical phenomena, which in turn addresses the main difficulty of case studies namely "handling the interrelatedness of various elements in the research work" (Dubois and Gadde, 2002, p.555). Systematic combining is a non-linear process meaning that theory and reality is continuously and simultaneously developing (Dubois and Gadde, 2002). According to Dubois and Gadde (2002) this approach is especially useful for development of new theories based on case studies. Systematic combining has been used in this thesis for the development of a framework which incorporates big data in forecasting, based on a case study at Volvo Group (VG). In figure 1 the basic principles of systematic combining can be viewed, where Dubois and Gadde (2002) describes matching as a process of moving between the theoretical framework and the empirical world in order to further analyze the development of new theories. According to Dubois and Gadde (2002), direction and redirection refers to the influences received from a cases study and its effects on the developed theory.

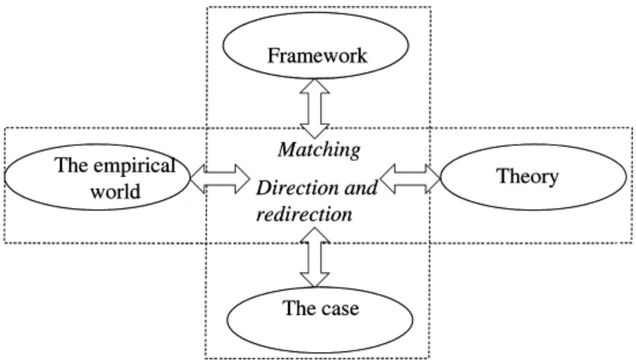

FIGURE 1: AN ILLUSTRATION OF THE FUNDAMENTAL PRINCIPLES IN SYSTEMATIC COMBINING (DUBOIS AND GADDE, 2002: P.555)

Framework

A theoretical framework was developed continuously throughout the thesis work and findings in the empirical data have expanded the literature study. Consequently this is in line with the following statement regarding purpose of literature reviews, "Selecting, reading and evaluating literature is an ongoing core activity of researchers that is usually carried out routinely and intuitively." (Seuring and Gold, 2012, p.545). According to Securing and Gold (2012) the aim of a literature review is to map, consolidate and evaluate the current state of knowledge in a specific field. An initial literature review was conducted in order to facilitate the collection of empirical data. This initial review focused on areas relevant to the purpose and research questions and was deemed necessary in order to give a solid foundation of knowledge needed for conducting the study and gather the empirical data.

The case and empirical world

According to Yin (2013) a case study is considered to be an appropriate way to answer "how "and "why" research questions. Furthermore, Yin (2013) states that it is appropriate when studying contemporary events and the rapid increase in available data can be considered as such. The study can be classified as an embedded single study which is aligned with the theory presented by Yin (2013). A case study at VG within their material management (MM) department was conducted as a basis for understanding correlations between available big data in an automotive aftermarket supply chain and its usage in forecasting, thus it can be described as the empirical world. A single case study can be used since VG can be seen as a typical and representative case as recommended by Yin (2013).

Theory

Based on the findings from the empirical world of the case study and the theoretical framework, a holistic theory of incorporating big data into forecasting was created. The theory is described as a framework which uses a classification commonly found in concurrent research. It is based on two dimensions, both of which are considered equally significant. The first dimension being the sophistication level of translating big data into a demand, which was developed through understanding which data is created in the aftermarket supply chain of an automotive company, such as VG. By combining the available data and theoretical framework a formula for translating big data into a demand was created. The second dimension describes different supply chain visibility levels for integrating big data in the planning process of spare parts. Through understanding and analyzing the current planning processes at an automotive company, such as VG, and comparing with the theoretical framework, this dimension could be developed. Combining both dimensions gave rise to the framework, which describes maturity stages in the focal company for incorporating big data into forecasting.

2.2 DATA COLLECTION

By using multiple sources for data collection it is possible to reveal unknown aspects or dimensions in case studies, thus complementing new discoveries (Dubois and Gadde, 2002). Furthermore, multiple sources give more weight to the case study. This thesis is based on a mixed methods approach, combining quantitative and qualitative techniques for data collection. This is also supported by Barton and Court (2012) that claims that a hypothesis of relevant data to study is needed in order to not be stuck in endless research. All of the empirical data was collected through VG. Furthermore, a literature review was conducted continuously throughout the thesis work. These data sources will now be described more in-depth.

2.2.1 LITERATURE REVIEW

Continuously throughout the thesis work literature was collected through mainly electronic means using the two databases Summon and Google Scholar. A combination of initial search words used in this thesis include: forecasting, big data, spare parts, aftermarket, supply chain visibility, correlation analysis and inventory management. Literature which was found to be relevant to the thesis work was also used as a means of backtracking in order to find additional literature. This method was found to be surprisingly effective and facilitated the credibility of the studied literature. Because it was identified that common sources was used in the studied literature. The literary sources used in this thesis were, in ranking order of frequency: journal articles, magazine articles, reports and books. In parallel to the continuous literature review all findings was scrutinized as a means of validation, this included among others investigating publication area, occurring citations and the trustworthiness of used sources. All literary sources have been denoted throughout the report with the Harvard referencing style and can be viewed in the reference list.

2.2.2 EMPIRICAL DATA

Empirical data was collected through four of the sources described by Yin (2013): (1) study of internal documentation and data, (2) direct observation of daily work, (3) participant observation and (4) interviews with relevant employees,

(1) Internal documents and data

Internal documentation refers to employee documents, internal company reports and records, e-mail correspondence (Yin, 2013). This was supplied or created through the development of this thesis. Data was also extracted from the company's internal databases and refers to any data that was correlated to facilitate the improvement of forecasting, this includes telematics data regarding trucks, sales data and forecasting data, such as historical demand. Additionally, certain documents was gathered internally from the company's web pages to support the understanding of the general vision of the company and the current internal research projects, sources could be documents such as annual reports and from the news segment.

(2) Direct observations

Observation is described by Bryman and Bell (2015) as a way to systematically observe and record of special phenomenon. The observation is to be performed such as where the researcher is taking part of some tasks that is done by the staff (Marshall and Rossman, 2014). Yin (2013) means that direct observations is a natural part of a case study since it takes place in a real life setting and are a way to discover behavior and phenomena and therefore serves as a evidence in the case study. The direct observations in this thesis was performed through spectating regular meetings that are part of the department in which this study was conducted, the direct observations include daily activity and operational tasks, weekly board meetings, continuous improvement meetings and meetings in sales and operations planning (S&OP) as well as field visits.

(3) Participant observation

Yin (2013) explains that participant observation is a special mode in which the observers take roles within the case study and participate in events. This study involves certain observation where the participants actively working to enhance the chance of using data within the supply chain in order to improve forecasting.

(4) Interviews

Yin (2013) states that one of the most important contribution source to a case study is achieved through interviews. The interviews are overall an essential source for case studies and contribute to evidence about human affairs and behaviors. Bryman and Bell (2015) emphasize that qualitative interviews offer a structure which maximizes reliability and validity, because it gives the interviewee the possibility of expressing their own view of the subject without an external influence driving the opinion. Two different ways of structuring interviews was used in this thesis as a form of qualitative data collection. They are denoted as semi-structured and unstructured interviews. The latter is based on a specific topic without order or amount of questions; this gives room for more flexibility where the interviewer can follow up the topic depending on how the interview evolves (Bryman and Bell, 2015). This approach can be used in order to facilitate the emergence of new theories and ideas based on the interviewee's knowledge, thus the interviewee sets the agenda and the interviewer listens and occasionally asks follow up questions (Corbin and Morse, 2003). Interviews with a semi-structured approach are described by Bryman and Bell (2015) as a way to cover a specific topic but where there are no specified alternatives for the question available. The interviewer structures the interview with an agenda based on specific questions regarding the subject in which the interviewee answers in their own accord (Corbin and Morse, 2003).

All interviews were conducted with employees at VG at different levels and departments, the choice between using a semi-structured or an unstructured approach depended on the goal of the interview. Interviews with a semi-structured approach was used in situations where an understanding of planning processes related forecasting of spare parts and the aftermarket supply chain within the organization was sought after. This approach facilitated the need of clarification regarding business terms and definitions used within the organization related to the discussed subject. Unstructured interviews were used in situations when the goal was to gain new insights and possible leads. Thus the interviews with such a characteristic created a contact network which was used in order to collect quantitative data. Due to the organizational structure of VG, which is characterized by silos that work independently of each other, a snowball sampling method was used. According to Bryman and Bell (2015) the snowball method, manifested in this thesis, is a technique where the interviewee's professional network is exploited in order to establish contact with new potential interviewees.

Within VG interviews was held with employees at Volvo Group Truck Operations (GTO), Volvo Group Truck Technology (GTT), Group IT and sales departments. In total there where 23 interviews with 25 participants from the mentioned subsidiary bodies, table 1 shows a summary of the participants. A complete list of the conducted interviews can be found in Appendix B which includes their title, department and the topic which was discussed. Additionally, in Appendix A there is a full interview guide which was used for the semi-structured interviews. In all interviews there were notes taken and several was also recorded for clarification purposes. Furthermore, a significant part of the unstructured interviews was conducted through the internet, which limited the possibility of recording. After each interview notes was transcribed, where extra focus lied in not mixing the interviewer's interpretation with the actual answers given.

TABLE 1: SHOWING THE NUMBER OF PARTICIPANTS IN ALL CONDUCTED INTERVIEWS FROM DIFFERENT DEPARTMENTS.

	GTO	GTT	GroupIT	Sales
Number of participants	7	6	2	10

2.3 RELIABILITY AND VALIDITY

In order to ensure high research quality Bryman and Bell (2015) suggests that researchers must evaluate the quality of the research outcome. Additionally, the authors state that the appropriate way to determine the quality of research studies is based on two criteria; reliability and validity. These are commonly used for quantitative research and can also be used for qualitative research while some debate that qualitative research should be evaluated differently (Bryman and Bell, 2015). However, based on this thesis' chosen research strategy it is appropriate to use both criteria as tools for assessing the quality of the study.

2.3.1 RELIABILITY

Reliability is a measurement which can be translated as replicability or credibility. For qualitative research reliability can be divided as internal and external (Bryman and Bell, 2015). The external refers to the probability that another research could mimic the study. More specifically that the ability to achieve the same result if a different study were to take place under similar circumstances, environment and setup. Additionally, internal reliability addresses the consistency of the findings, indicating if participants in the study identify the same phenomenon or problems. There are several techniques to increase reliability of a research study according to Riege (2003), for example to use multiple researchers, ensuring multiple findings or conducting a more thorough documentation.

This study is based on both qualitative and quantitative research methods. Interviews were conducted in order to gather qualitative empirical data. By thoroughly taking notes, recording, transcribing and reviewing each interview by multiple researchers, internal reliability could be ensured. Additionally, a consistent cross-checking of the interviewer's outtake with the participants further ensures and strengthens the findings, thus minimizing interpretational discrepancies which add to the internal reliability. The quantitative empirical data which was gathered is based on large data-sets received from connected vehicles, sales data and historic demand data. Its internal reliability was confirmed by stakeholders who are directly responsible for the respective databases which contain the data-set. A validation of the data was executed in coordination with relevant employees at VG in order to ensure its accuracy and quality.

2.3.1 VALIDITY

Validity is according to Bryman and Bell (2015) a measure of to what extent the observation, concepts and interviews are corresponding to reality. Validity can, similar to reliability, be divided as internal and external. The external refers to how relevant to findings are for the concept while the internal consider how the findings can be generalized for further research. Furthermore, Riege (2003) states that a research study's validity can be enhanced by using illustrative frameworks and diagrams or comparing findings with relevant literature or studies.

The findings of the thesis was based on the theoretical framework and the empirical data found at VG through a case study. By validating the findings with current planning processes at VG in collaboration with employees at the company the validity of the findings could be secured. The external validity of the findings was ensured through basing the developed theory on empirical findings and the theoretical framework. The developed theory, described as a framework, generalizes the findings from the empirical study and theoretical framework thus securing its usage in further research and ensures the internal validity. Considering that triangulation of data sources was used in regard to the empirical study increases the internal validity. This was done through reviewing the empirical findings with the studied literature presented in the theoretical framework.

3. THEORETICAL FRAMEWORK

This chapter describes the reviewed literature which is relevant in order to fulfill the purpose of the thesis. It provides a basis for understanding the main topics of the thesis, areas such as the aftermarket, spare parts, forecasting, big data and information flow are characterized and reviewed, providing a fundamental foundation of knowledge which is needed for analyzing the empirical findings and later interpreting the results.

3.1 THE AFTERMARKET

Selling of a new product does not mean that business opportunities or responsibilities ends, because regardless of how well crafted a product might be there will come a time in the product life-cycle where it will not conform to its design specifications thus creating a failure during usage (Cohen and Lee, 1990). This concept explains the aftermarket where services are offered in order to restore the product's original functionality.

3.1.1 THE AFTERMARKET POTENTIAL

The aftermarket is an opportunity which has been described as "the golden age of service" (Cohen et al., 2006; p129), which in business means that every manufacturer must take the opportunity to transform into being a service business. Likewise the automotive industry cannot consider the sale of a new product as the end in a business transaction but should instead see their relationship with their customers as long term and provide services during the whole product life-cycle (Cavalieri et al., 2007). According to Cohen et al. (2006) General Motors had more than 16 times more profit margin on their aftermarket in comparison to new product sales. The aftermarket has shown to be for many companies a larger source of both revenue and profit then the new product market, and many studies show that the profit margin is significantly higher in the aftermarket than for new products (Bartwal, et al. 2010; Bundschuh and Dezvane, 2003; Gaiardelli et al., 2007). Hence, the aftermarket creates new opportunities and has had an increasingly interest from management at manufacturing companies, since it brings competitive advantages and huge potential source of increasing profitability (Cohen et al., 2006). The aftermarket also enables the possibility to bond stronger relationships with the customer and therefore plays a role in increasing the customer satisfaction and company reputation, which becomes important during new product launches (Goffin, 1999; Goffin and New, 2001; Phelan et al., 2000).

3.1.2 THE AFTERMARKET SUPPLY CHAIN

According to Cohen et al. (2006) it's important for automotive companies to transform into service businesses but also to adopt its supply chain making it suitable for the aftermarket service. The main goal according to Chopra and Meindl (2013) for a supply chain is to maximize the surplus and fulfill customer's needs; this also applies for a service supply chain in the aftermarket. The supply chain is defined by Handfield and Nichols (2002) as the organization's activity and flow that is related to the transformation of goods from being raw material through to the final end customer; it also includes the monetary flow as well as the involved information exchange. While the management of supply chains is related to the relationship and activities among the actors of organizations (Golicic et al., 2002). Generally a supply chain can be divided into two trade-offs that need to be considered, responsiveness and efficiency (Chopra and Meindl, 2013). The different tradeoffs in the supply chain are categorized by Chopra and Meindl (2013) and are shown in table 2.

TABLE 2: A COMPARISON OF A RESPONSIVE AND EFFICIENT SUPPLY CHAIN (CHOPRA AND MEINDL, 2013, P 42)

Parameter	Efficient Supply Chains	Responsive Supply Chain
Primary goal	Supply demand at the lowest cost	Respond quickly to demand Create modularity to allow postponement of product
Product design strategy	Maximize performance at minimum product cost	Postponement of product differentiation
Pricing Strategy	Lower margins because price is a prime customer driver	Higher margins because price is not prime customer driver
Manufacturing Strategy	Lower cost though high Utilization	Maintain capacity flexibility to buffer with demand/supply uncertainty
Inventory Strategy	Minimize inventory to lower Cost	Maintain buffer inventory to deal with demand/supply uncertainty
Lead time strategy	Reduce, but not at the expense of cost	Reduce aggressively, even if the cost are significant
Supplier strategy	Select based on cost and Quality	Select based on speed, flexibility, and quality

Furthermore, one can make a distinction between the aftermarket supply chain and a manufacturing supply chain, where the difference according to Cohen et al. (2006) is that the former manages significantly more stock keeping units (SKU). A more thorough comparison of the two can be viewed in table 3.

TABLE 3: COMPARISON OF A MANUFACTURING AND AN AFTERMARKET SUPPLY CHAIN (COHEN ET AL., 2006: P.132).

Parameter	Manufacturing supply chain	Aftermarket supply chain
Nature of demand	Predictable, can be forecast	Always unpredictable, sporadic
Required response	Standard, can be scheduled	ASAP (same day or next)
Number of SKUs	Limited	15 to 20 times more
Product portfolio	Largely homogenous	Always heterogeneous
Delivery network	Depends on nature of product; multiple networks necessary	Single network, capable of delivering different service products
Inventory management aim	Maximize velocity of resources	Pre-position resources
Reverse logistics	Doesn't handle	Handles return, repair and disposal of failed components
Performance metric	Fill rate	Product availability (uptime)
Inventory turns (the more the better)	6 to 50 a year	1 to 4 a year

The aftermarket supply chain contains two significant characteristics which play an important role in its success, forecasting and information sharing. Cohen et al. (2006) mentions that companies view the aftermarket supply chain the same as they would the manufacturing supply chain, which then would lead to an incongruity between supply and demand, because of the inconsistent nature of the aftermarket demand. Additionally, forecasting for the aftermarket supply chain should not only be based on historic demand but instead also use input from a combination of sources within the focal company, such as sales data, end-customer data or economic indicators (Baudin, 2004). The second factor, information sharing, plays an important part in manufacturing supply chains, where information regarding production schedules and inventory levels are shared both up-and downstream. The main reason for this is to create visibility and counteract the bullwhip effect, which refers to a distortion of information that results in increasing variances in demand further upstream the supply chain (Lee et al., 1997). Information technology (IT) has had a significant impact on information sharing in regard to customer service; however it has been neglected in the aftermarket supply chain (Phelan et al., 2000).

The complexity and rough nature of an aftermarket supply chain makes most companies neglect its design and management and thus its performance suffers (Cohen et al., 2006). The authors further state that this is partly due to unpredictable and sporadic demand pattern, which requires responsiveness, the large number of SKUs and high product variety. Which is directly relatable and in line with the challenges in spare parts control presented by Huiskonen (2001), who identifies that difficulty in forecasting, high service requirements and large gaps between pricing of products, creates significant financial impact if stock-outs occur. This further argues for the need of tackling the challenges of designing and managing the aftermarket supply chain, which has to cover all products that are promised to the customer. Wanger and Lindeman (2008) emphasize, that spare parts is one of the most profitable sources of income for manufacturing companies.

3.2 SPARE PARTS

In the early 90s Petrović and Petrović (1992) stated that there was an increased interest in spare parts management which was attributed to two factors:

1. Reliability is not a certainty meaning that failures always occur and having a selection of spare parts on-hand near the customer is important for high uptime and operation.

2. Investments in spare parts, both in regard to stock-value and IT-systems, can be a significantly large portion of the products capital value.

Additionally, Slater (2010) states that spare parts availability is critical in maximizing uptime. Recent studies have shown that the complex material control for spare parts lies in the highly unpredictable demand (Cohen et al. 2006, Jouni et al. 2011). It is also emphasized by researchers that the contributing factors is partly due to the fact that a spare parts supply chain requires responsiveness, high number of SKUs and high variety of products (Cavalieri et al.,2008; Cohen et al., 2006). The number of spare parts in an aftermarket supply chain can be more than 20 times than for a manufacturing supply chain (Cavalieri et al., 2008). The underlying reason for that can be explained by the short life span and the commitment that companies undertake for their customers in providing spare parts during the entire product life-cycle (Cohen et al., 2006). Kennedy et al. (2002) explain that the inventory levels main reason is to protect against irregularities in the demand and by reducing the uncertainty companies can maintain a more smooth material flow with lower inventories. Additionally, Jouni et al. (2011) state that classification of spare parts is a way to reduce the complexity of the material flow and contributes to making the whole chain more efficient. This is a simpler way to manage the parts instead of handling all the spare part numbers.

Spare parts classification

The sheer number of spare parts, which often is significantly larger than production parts in the automotive industry, increases the complexity of the aftermarket supply chain; implementing different classification for similar spare parts creates a much more efficient management than handling spare part numbers (Jouni et al. 2011). In order to help better determine forecasts or stock control decisions it is vital to have a classification, mainly because spare parts can have a highly varied costs or demand pattern (Boylan and Syntetos, 2008). According to a study by Bacchetti and Saccani (2012) the most popular classification criteria is using the part cost, either as a unit or inventory cost. Additionally, as mentioned by Bacchetti and Saccani (2012), there are other types of common classification criteria such as using demand volume or value. Furthermore, the authors mention supply characteristics as a classification; where one could use replenishment lead-time, availability at supplier or the risk of non-supply.

Many researchers, such as Lawrenson (1986), Kennedy et al. (2002), Romeijnders et al. (2012) have made a distinction in two different types of maintenance. The first is planned maintenance which is done to preventive failure and is generally based on fixed time intervals, related to usage or the condition of the product. Since the demand is directly linked to what in many cases are fixed parameters, the demand of a spare part needed for a preventive maintenance is, in theory, easy to predict. The logic is that the closer the interval is between the maintenance the higher the demand will be. Accurate prediction of demand gives the company opportunity to deliver it efficiently. The second type of maintenance is called corrective maintenance, this maintenance is scheduled when a part has already failed or is close to failure. This type of failure is harder to predict since breakdowns normally occur at what is perceived as random time intervals.

Classification can be done in order to reduce the complexity of the supply chain and therefore be able to manage spare parts more effectively. This enables management of spare parts depending on group level instead of spare part number level (Jouni et al., 2011). The classification can aid in determine the inventory level needed as well as choosing the forecasting approach appropriate for each classification (Boylan and Syntetos, 2008). A widespread classification is based on the Pareto-principle which separate the spare parts depending on the price of the item and volume as well as additional characteristics that could be considered important (Huiskonen, 2001). Boylan and Syntetos (2008) argue that of these additional characteristics could be set to the criticality of the spare part.

3.3 SUPPLY CHAIN VISIBILITY

The operational activities of the aftermarket can easily be assumed to be a hassle free business area and that no more concern is required from managers (Aberdeen Group, 2006). Reality has showed quite the opposite, in a study of 150 companies, even the ones that are above the average still has to get rid of the silos that they are working in and get a more holistic view (Cavalieri et al., 2008; Aberdeen Group, 2006). Robinson (2014) emphasize that the digital requirement will increase in order to achieve higher service level in the aftermarket. This requires an increase in information and communication technology, allowing manufacturers to receive updated data in real time. This is supported by Golicic et al. (2002) who states that members of supply chain can use real time data via internet in order to be more responsive.

Moreover, in the automotive industry, data received through connected vehicles gives according to Frowein et al. (2014) new opportunities such as to provide interactive maintenance and real time diagnostics. The information collected between actors in a supply chain can be explained as the supply chain visibility. Academia has done several attempts to approach a concept for the supply chain visibility over the years in order to explain it. One of the attempt made by McIntire (2014) explain the supply chain visibility as the process that involves collecting data related to supply chain, integrate it and extract intelligence in order to make decisions. Another definition by Barratt and Barratt (2011) is to what extent “actors within a supply chain have access to or share information which they consider as key or useful to their operations”. An effective supply chain management lies in achieving supply chain visibility and that it is the greatest difficulty for companies. Companies still lack visibility and which is key driver to reduce complexity in an increasingly complex global operation (Heaney, 2013).

The visibilities main purpose is to enhance the decision making. This is done by “simplifying, accelerating, reduce the chance of failure or improving the completeness” (McIntire, 2014, p.12). By simplifying one can reduce the complexity in the supply chain which is often cited as a way to increase the value of supply chain management. Supply chain visibility can in many cases be viewed as a series of activities that takes place over time. By increasing the visibility the management can predict upcoming events in the near future, which gives room for better planning. Cassivi et al. (2005) states that the visibility can be achieved by sharing forecasts, schedules and production capacity with all supply chain stakeholders. Further Golicic et al. (2002) explains that visibility contains two components. These components are interaction and market access. Interaction is based on businesses today that are relying more on system connection with both supplier and customers. This creates an opportunity to remove the barriers among the supply chain actors and thrives on opportunity for supply chain participants to communicate and share information (Golicic et al., 2002). The second component, market access, gives companies the ability to access customer's data and understand their needs. Increasingly connected supply chains through the use of IT and the internet creates large amounts of data, which gives the focal company possibility to better satisfy their customers' needs.

The visibility related to downstream activities of the supply chain is mainly where the management is interested in making the supply chain more effective, giving higher customer satisfaction and increase agility (McIntire, 2014). There are many ways the information can be shared downstream, one example is the retail system implemented by Walmart in 1990 in order to get point of sales (POS) data and the result of this implementation is significant and it's a great example of improved visibility (McIntire, 2014). Other studies has also shown that visibility improves the supply chain, the best in class compared to the worst had 23 percent more on time deliveries and 11 percent better landed cost (Aberdeen Group, 2012). These companies were 2.5 times more likely to have visibility to events related to end customers. Cohen et al. (2006) states that an agile supply chain has to synchronize it supply with the demand since it is more market oriented. This synchronization is achieved by integrating the internal functions and also other actors within the supply chain.

3.3.1 PREREQUISITES FOR SUPPLY CHAIN VISIBILITY

The steps shown in figure 2 describe a required process which is needed in order to achieve supply chain visibility. The process starts with data collection, which is how data is collected and saved in addition to what system is used. The following step is the integration of data and relates to the correlation and interconnection of data which has been collected; it could for example be used for verification or identification. This integration leads to the third step, the ability to create intelligence from the data. This is the step where the data is transformed into something meaningful for the user. The final step of this process is where the intelligent data is used to interrupt or support the decision making process which is performed by the user. (McIntire, 2014)

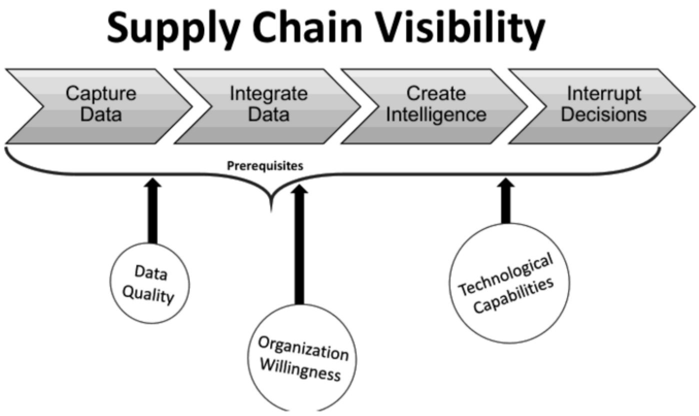

FIGURE 2: THE PROCESS OF HANDLING INFORMATION WITHIN AN ORGANIZATION. (MCINTIRE, 2014)

According to McIntire (2014) academic research suggests that two factors are reviewed in order to achieve a successful supply chain visibility; the quality of the data and the organizational commitment. The first, data quality (also called information quality), is related to what extent the data meets the need of the organization. In practice this can be quantified as; accuracy, trust, timeliness and usability and the quality is directly linked to the success of the supply chain visibility (McIntire, 2014). The latter factor, organizational commitment, is related to the willingness of the participants to make necessary changes needed in order to achieve the desired success. The reason is that the focal company must make large changes within the organization and this naturally scares people involved who are affected by the change. McIntire (2014) identifies four steps in the process of handling information in an organization, visualization can be seen in figure 1 and the steps are now to be reviewed.

3.3.2 VALIDATING SUPPLY CHAIN VISIBILITY

In order to validate the different sub-processes McIntire (2014) suggest using a developed matrix as a scorecard. Which uses scores from four different matrices where each of them represent the performance of the four steps within supply chain visibility process shown in figure 1. The four steps are called: (1) sensitivity, (2) accessibility, (3) intelligence and (4) decision-relevance. All of which is presented below, in addition a final scorecard from McIntire (2014) is also exemplified.

(1) Sensitivity
The first matrix (see table 4) is called "sensitivity" and represents the performance of the first step, capturing data. The source of demand management could be "historical data, sales projections, promotion plans, corporate objectives, market share data, trade inventory market research and new categories of growth" (Croxton et al., 2001, p.18-19). Croxton et al. (2001) further if vendor management inventory is implemented the company receive data direct from the end-costumer. High score means that the process in capturing data is very successfully.

TABLE 4: THE "SENSITIVITY" MATRIX SCORECARD RELATED TO THE FIRST STEP IN THE PROCESS OF ACHIEVING SUPPLY CHAIN VISIBILITY. (MCINTIRE, 2014)

Score	Description
0	No data is captured to support the target business decision.
1	Some relevant data is captured, but it is incomplete.
2	All data is captured but the accuracy of the data is unknown or known to be low.
3	Data is complete and consistently biased (i.e. low quality but predictable).
4	All data needed to support the decision is captured, complete, consistent and measurably high in accuracy.

(2) Accessibility
The second matrix (see table 5), called "accessibility", is directly linked to the second step, integrate data, and a high score indicated that the user can navigate and move through the nodes or objects with low cost and time. Having a cross functional integration of communication within companies is a prerequisite to achieve successful operation and to enhance the business outcome (Croxton et al., 2001).

TABLE 5: THE "ACCESSIBILITY" MATRIX SCORECARD RELATED TO THE SECOND STEP IN THE PROCESS OF ACHIEVING SUPPLY CHAIN VISIBILITY. (MCINTIRE, 2014)

Score	Description

0	Data remains in the capturing systems with no attempt to integrate the data for later use.
1	Data remains in the capturing systems, but processes allow them to be manually integrated for ad-hoc tasks.
2	The solution integrates all the decision-relevant data, but not all of it is retrievable by the decision maker.
3	Data is integrated and available to the decision maker, but not using methods they prefer.
4	All relevant data is integrated and accessible by any relevant path the decision maker could use.
5	All relevant data is integrated, accessible and the approach to integrating data is easily adaptable.
6	All relevant data is integrated, accessible and the integration approach is self-updating when confronting new data types or sources.

(3) Intelligence

The third matrix (see table 6), defined as "intelligence", represent the step in which intelligence is created. This is the hardest to quantify but generally a high score mean that the data is easy to understand and to identify if an interruption should be made. Furthermore, it could be viewed as the ability to learn and develop through performance feedback (McIntire, 2014). When the data is in hand the forecast can be devolved and if the source comes from the costumer it could give critical information expected demand. Tracking and analyzing the forecast accuracy could be used as feedback to fine-tune the method (Croxton et al., 2001).

TABLE 6: THE "INTELLIGENCE" MATRIX SCORECARD RELATED TO THE THIRD STEP IN THE PROCESS OF ACHIEVING SUPPLY CHAIN VISIBILITY. (MCINTIRE, 2014)

Score	Description
0	There is no automated recognition from the solution that a business decision is needed.
1	Sometimes there is recognition from the solution that a business decision is needed.
2	The solution always knows that the business decision is needed.
3	The solution's approach to recognizing the need for a business decision is easily updated by users.
4	The solution's approach to recognizing the need for a business decision is self-updating.

(4) Decision-relevance

The last matrix (see table 7) is ""Decision-Relevance" and is the performance of the last step, a high score would indicate that the decision is specified autonomously and does not need any further assistance from user.

TABLE 7: THE "DECISION-RELEVANCE" MATRIX SCORECARD RELATED TO THE FOURTH STEP IN THE PROCESS OF ACHIEVING SUPPLY CHAIN VISIBILITY. (MCINTIRE, 2014)

Score	Description

0	The solution has no explicit input to this business decision.
1	The solution is a required information source for the decision maker. A user decides how and when to make the decision.
2	The solution is a required information source for the decision maker. The solution decides when the decision is taken and the user decides everything else.
3	The solution offers a set of action alternatives based on the event, or
4	narrows the selection down to a few, or
5	suggests one action, and
6	executes that suggestion if the human approves, or
7	allows the human a restricted time to veto before automatic execution, or
8	executes automatically, then necessarily inform the human, or
9	informs the human only if asked, or
10	the solution decides everything and acts autonomously, with no notice given to humans except for debugging.

Final scorecard

The final step after giving the supply chain visibility a score from these scoreboards is to specify the cost of implementing such solution. This includes all the steps and is not focused to any of the singular steps in the process of supply chain visibility. The final scoreboard could look something like table 8 shown below.

TABLE 8: THE FINAL MATRIX SCORECARD SHOWING A COMPARISON OF THE SOLUTIONS RELATED TO THE PROCESS OF ACHIEVING SUPPLY CHAIN VISIBILITY. (MCINTIRE, 2014)

Solution	Sensitivity	Accessibility	Intelligence	Decision Relevance	Fitness	Solution cost in Thousands of USD
A	2	2	3	1	33%	850
B	3	5	0	5	54%	1,120
C	1	3	2	3	38%	775

McIntire (2014) emphasizes that the impact of the supply chain visibility must be understood in order to be able to understand whether an investment is beneficial or not. The fitness relates to the potential improvement in supply chain operation which in turn improves the business outcome as seen in figure 3. If the improvement of business outcome covers the cost for implementing solution the investment is considered as profitable.

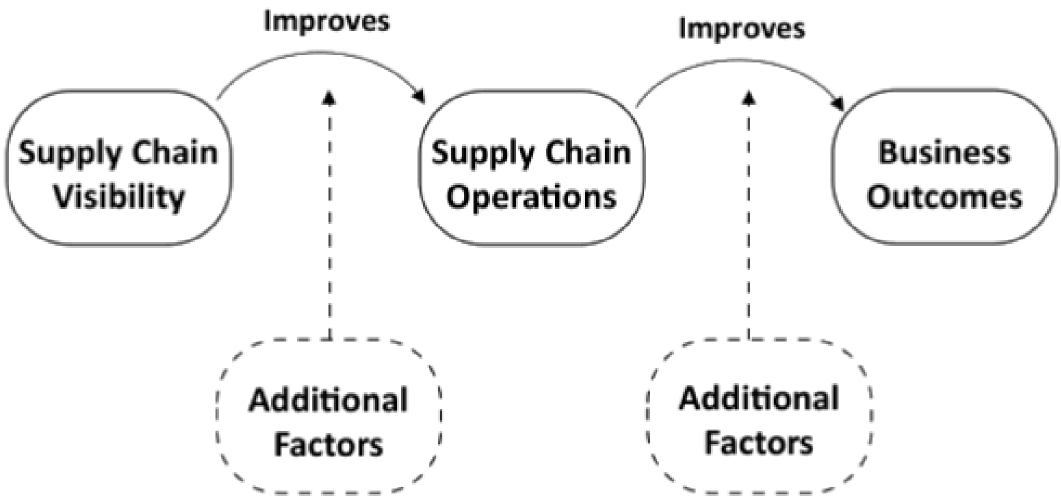

FIGURE 3: VISIBILITY RELATION TO THE OVERALL BUSINESS PERFORMANCE (MCINTIRE, 2014, P.35)

3.4 BIG DATA

As one of many concurrent buzzwords the big data concept is not only a trend but also a concept which offers significant opportunities to change modern business models and day-to-day decisions (Waller and Fawcett, 2013). A study conducted by IBM and Oxford University showed that companies today have different views in regard to defining what big data is (Schroeck et al. 2012). The respondents from the study defined big data as everything from a greater scope of information to social media data.

3.4.1 BIG DATA DEFINITION

According to Shi (2014) the international scholars at Xiangshan Science Conference defined big data as "a collection of data with complexity, diversity, heterogeneity, and high potential value that are difficult to process and analyze in reasonable time" (Shi, 2014, p. 6). And similarly Manyika et al. (2011) defined big data as "... datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze" (Manyika et al., 2011, p.1). However, researchers generally agree on characterizing big data according to the three V's; volume, velocity and variety (Zikopoulos, et al., 2012).

Volume, which refers to the increasing amount of storage capacity related to the data. During 2012 it has been estimated that 2.5 exabyte of data was created every day (McAfee et al., 2012). Putting this in relation to the total amount of data which was stored in global computer networks in 2000, about 200 exabytes, gives an indication of what the future of how ever increasing technological advancements might strain the global IT infrastructure; in 2020 it is estimated that total amount of data will increase to 35,000 exabytes (Zikopoulos et al., 2012).

Velocity refers to partly the short time in which the data is valuable, but mostly to the swiftness at which the large amounts of data can flow (Zikopoulos, et al., 2012). According to McAfee et al. (2012) the rate at which data is collected and stored is more important and has a more significant impact in later analysis. Because it facilitates a more agile strategy and thus companies can outcompete their competitors.

Variety is regarded as the differentiated sources of which data can be extracted. It also refers to the varying ways data can be structured or unstructured, depending on the source it can be increasingly complex to analyze data for a specific task because of the way the data is collected (Zikopoulos et al., 2012).

3.4.2 IMPACT OF BIG DATA

A UN initiative called Global pulse is harvesting big data from social networks in order identify early warning signs and thus creating proactive assistance programs, for example understanding health trends in a certain population and thus stopping outbreaks (World Economic Forum, 2012). For the private sector big data will impact not only the increase in customer value that can be offered through new and innovative solutions but also significantly improve internal processes (Lohr, 2012; Accenture, 2014). However, in today's business setting companies must tackle the challenge of having access to the right information at the right time, in order to achieve competitiveness (Uschold and Gruninger, 2004). Research has shown that the available data will continue grow in the foreseeable future and organizations that aren't willing to embrace these challenges and get the skills required, will find themselves lacking in competitiveness (Hassani and Silva, 2015; Accenture, 2014).

Published research shows that one-third of all top-field companies that use big data for managerial decisions, are six percent more profitable than their competitors (Mcafee et al., 2012; Lohr, 2012). Because managers can know radically more about their own business, which can directly improve the performance of the organization. The director of research at Google simply described it as "we don't have better algorithms, we just have more data" (Mcafee et al., 2012, p. 63).

One of the biggest challenges for implementing big data in forecasting is the unaware hindrance created by concurrent scientist, researchers and statisticians who are inveterate in traditional statistical techniques, which are used in order to obtain precise forecasts (Hassani and Silvia, 2015; Arribas-Bel 2014). In accordance with this Arribas-Bel (2014) further states that a major challenge for the future of big data in forecasting is the advanced skills which are needed in order to understand correlations and implement new models. According to Shi (2014) the fact that big data uses real-time information makes it increasingly important to use techniques which can transform unstructured data into structured data.

Nonetheless in today's increasingly connected and globalized world, big data can be viewed as something of the past (Hassani and Silvia, 2015). The authors mean that the future is more concerned about making use of the collected data because the large amounts of information available in today's computer networks are not being utilized to its fullest potential. Furthermore, Hassani and Silvia (2015) explain that companies that aim to increase their competitiveness in the future must invest in the development of tools which exploits big data, today. One of these tools, as mentioned by Hassani and Silvia (2015) is the potential for greatly improving forecasts. As said by Ton (2013) forecasting and big data goes hand in hand and can be improved by using real time forecasting by exploiting the increasingly available data and Tucker (2013) believed that big data soon can be used to predict every organizational move.

3.5 FORECASTING

The intermittent demand, normally cased in the aftermarket, is very difficult to predict. The way to forecast spare parts with a lumpy demand has been on debate in academia for a long time and started with Croston (1972) that proved that single exponential smoothing is not the right way to go when forecasting demand with irregular patterns. The reason for such pattern is the huge number of SKUs which contributes to a large uncertainty in the demand (Cohen et al., 2006; Jouni et al., 2011). Cohen et al. (2006) emphasize that many companies are using similar approach when forecasting spare parts as for manufacturers supply chain which cause a mismatch in the supply and demand due to the sporadic nature in demand. In the aftermarket spare parts are distributed based on a probability that has to be forecasted due to the expectancy of breakdowns. Companies should now manage the material flow in a way that mitigate the risk and not only match the forecast. The result is that companies will face incongruity between supply and demand (Cohen et al., 2006). Baudin (2004) explain that companies should not forecast solely on the use of algorithms based on historic demand and instead use the input from customers and sales as well as leading economic indicator. However, spare parts generally lack reliable information in order to predict when failure occurs and especially when products newly enter the market (Kennedy et al., 2002). Cassivi et al. (2005) suggests that business partners within the supply chain should share forecasts. He further elaborates that not only forecasting but also schedules and production capacity and other types of accurate information should be visible for the entire supply chain.

3.5.1 FORECASTING METHODS

Cavalieri et al. (2008) describes two different type of forecasting methods dependent on the available data for spare parts. Firstly, time-series based forecasting is a technique that uses the historical demand in order to forecast the future demand, which is the most common practice, used in industry (Cavalieri et al., 2008). The method does not require any information of the operation condition and the product base. The approach is less appropriate when the demand is very low or when introducing new products which lacks in the amount of data that can be used to determine future demand (Cavalieri et al., 2008).

The second forecasting method, called reliability based forecasting, refers to forecasting which is based on the amount of installations and the technical operating condition (Cavalieri et al., 2008). The main goal is to determine the demand by using reliable information of the spare parts and accumulate the operating times (Cavalieri et al., 2008). The estimation could be by collecting failure rates from the different types of spare part items. Petrović and Petrović (1992) suggest that the forecasting could be calculated based on the failure rate. This in turn requires data regarding the amount of products to maintain, the amount of spare parts inside each product, the intensity of the operation and parts prediction failure rate. The failure rate can be determined by using life data analysis of the historical failures or reliability tests and other data could be accessed through data banks. This method is according to Cavalieri et al. (2008) particularly good when the spare parts are new and there is limited historical demand. The drawbacks might be that the lack in accurate data of failure rates and conditions. Another fall pit could be the innovations rates of complainants which make it hard to estimate the failure rates. A flowchart showing practical methods for using either time-series based or reliability based forecasting can be seen in figure 4.

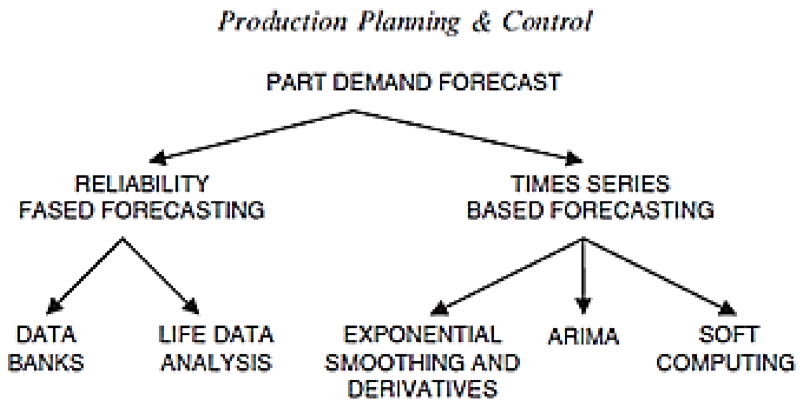

FIGURE 4: TWO METHODS FOR FORECASTING SPARE PARTS, WITH PRACTICAL TOOLS FOR EXECUTION. (CAVALIERI ET AL., 2008)

Life data analysis is another way that could be used to understand the demand of an product, by using continuous distribution in order to determine the probability of failure rates depending on how much the item has been used, for example hours, month, cycles (Nelson, 2005). This could "be used for mostly anything to determine the life such as products, materials, people television programs, and many other things" (Nelson, 2005, p.16). One distribution that is described by Nelson (2005) is the normal distribution and is appropriate when the value is a sum of a large number of random variables. Normal distribution is not naturally always the best option to use for life data analysis, Nelson (2005) suggests instead Weibull distribution for most cases. But Weibull would require much more technical understanding and tests for each spare part and in order to determine the shape.

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

EQUATION 1: EQUATION FOR GAUSSIAN DISTRIBUTION, ALSO KNOWN AS NORMAL DISTRIBUTION.

Where *F(x)* can be seen as the probability of failure, *x* is the chosen parameter, for example hours, months, cycles rates etc. μ is the mean and σ is the standard deviation. Both of which must be of the same unit as *x.*

From He and Tao (2014) it can be viewed that predicting demand, i.e. forecasting, is done through multiple sets of methods which has been research by scholars extensively. Furthermore, regression models can be used in order to predict outcomes based on intrinsic relationship between different variables (He and Tao, 2014). According to Puntanen (2010) regression models are based on understanding relationship between variables, it dates back to 1805 when it was used for the astronomical observation. This type of analysis is often found in research related to health for understanding different behaviors in relation to a sickness. Regression can be applied in numeral areas such as "medicine, biology, agriculture, economics, engineering, sociology, geology, etc." (Puntanen, 2010, p.4). One of the main goals that regression is used for is to predict the outcome of one variable by studying one or several other. It has for example been used in several studies to study the demand for water, where several social-economic and climate factors taken into consider in order to estimate the consumption in a particular area (He and Tao, 2014). Regression models are in principle a way to measure the correlation to identify similar or dissimilar relation between factors. The formula for calculating variables dependent on a set of variables with regression is the following:

$$Y(t) = X_1(t) + \cdots + X_n(t) + \varepsilon$$

EQUATION 2: MULTIVARIABLE REGRESSION EQUATION.

Y is the dependent variable and X is the independent variable, meaning that Y is dependent on X. The dependencies are similar to a multi-variable function with the additional variable ε that represents the random error (Puntanen, 2010).

3.5.2 FORECAST ACCURACY

There have been many suggestions on how forecast accuracy should be measured (Hyndman and Koehler, 2006). One of the most common is Mean Absolute Percentage Error (MAPE) and the benefit of using percentage error is that it is scale-independent and is therefore used when comparing different forecast accuracy values and data sets (Hyndman and Koehler, 2006). The percentage error is given by equation 3:

$$P_t = 100 * \frac{e_t}{Y_t}$$

EQUATION 3: CALCULATION METHOD FOR DETERMINING PERCENTAGE ERROR OF FORECASTS.

The variable e_t is the forecast error and is determined by subtracting the observed result Y_t from the forecast F_t. Since each of the variables is time based it's possible to study it over a period of time with many forecasts and observations. After determining the percentage error P_t, the MAPE can be set to mean ($|P_t|$) (Hyndman and Koehler, 2006).

3.6 CONCLUSION OF LITERATURE REVIEW

Based on the conducted literature study it can be suggested that the end-customer demand for spare parts could be estimated through use of multivariable regression and life data analysis. The formula is based on a multivariable regression model where the independent variable is determined by estimating the probability for a spare part to break. The formula which describes the demand for a specific time can be seen below.

$$D(t) = K\,(P_1(x_1(t)) + \cdots + P_n(x_n(t))) + \varepsilon$$

EQUATION 4: TOTAL END CUSTOMER DEMAND BASED ON LIFE DATA ANALYSIS.

The dependent variable *D(t)* represent the actual customer demand for any given time. Where *x(t)* is life data and refers to a measure that has any correlation with probability of failure. The data could be anything such as mileage, runtime, cycles, life length etc. The independent variable P represent the probability that a failure occur and could for example be built up by any distribution of probability, which is explained by Nelson (2005) as a way to determine when a breakdown occurs, the probability has a value between zero and one. If the company has what can be called as "complete data" the failure is already known and the P would therefore be either zero or one for the time given. All the different probabilities for each spare part existing on the market, i.e. the population, are summed up as a way to describe the total demand. K is any bias that represents the difference in the amount of failure occurring in the market and the demand; this could for example be affected by customer loyalty. The variable ε describes the nature of uncertainty and any random factor that impacts the demand.

The formula describes a concept where the actual demand of a spare part should in theory have a strong correlation with the probability that it breaks. This is exemplified by a simple description; if there are one hundred articles that have ten percent probability that a break will occur within one month, and customer loyalty is 70 percent, the demand would then for the upcoming month could be estimated to seven. If this analysis was done with an accurate probability for the whole population the result would be a forecast of the total demand for one particular spare part model. The theory of regression which is used to determine the stated equation explains how one should interpret uncertainties as well as how variables correlate with each other. The uncertainties occurred because the total customer demand is theoretical and varies in practice by randomness. That is achieved by includes the variable ε that explains randomness which could be interpret to explain the random pattern given by the aggregated probability of failure. Consequently a regression model gives the possibility of correlating external data to a specific demand, thus one can create demand patterns based on data other than historical demand.

4. EMPIRICAL DATA

In this section all of the empirical findings will be presented. It revolves around general information about VG as a global automotive manufacturer, the construction of their aftermarket supply chain, how planning processes are executed in order to forecast demand for spare parts and which data parameters created within the internal supply chain that, according to the company, has a potential to increase the accuracy of forecasting the demand for spare parts.

4.1 VOLVO GROUP INFORMATION

The Volvo Group is one of the world's leading manufacturers of trucks, buses, construction equipment and marine and industrial engines. Additionally, VG provides its customers with entire solutions for financing and service. Its headquarters is located in Gothenburg, Sweden and currently VG employs roughly 100,000 people around the globe. VGs production facilities are located in 18 different countries and they sell their products to a global market reaching more than 190 countries. In 2015 VGs net sales amounted to approximately 313 billion SEK and it is a publicly held company, listed on the Nasdaq Stockholm Exchange as AB Volvo. The company was founded in 1926 as a subsidiary body of AB SKF, a Swedish manufacturer of bearings. It was founded on the basis of being a car manufacturer but over the years expanded its business areas into more heavy duty vehicles and engines. In 1999 the VG parted with the Volvo Car brand, which was sold to the Ford Motor Company, and solely focused on their current business areas. Today VG provides a wide range of brands and products for customers; this has been done through mergers, acquisitions and joint ventures. Their current brands are: Trucks, Buses, Construction Equipment and Marine Engines. All brands offered are distinguished as individual business areas, which can be viewed in figure 3. Their current offering includes sales, services and aftermarket products for all of their brands.

FIGURE 5: ALL BRANDS OFFERED BY VG. (VOLVO GROUP, 2016B)

The organizational structure of VG was changed as of Mars 2016, with a new focus of making each brand responsible for their own sales function instead of being centralized. An overview chart of the business areas and divisions within the VG organization can be viewed in figure 4.

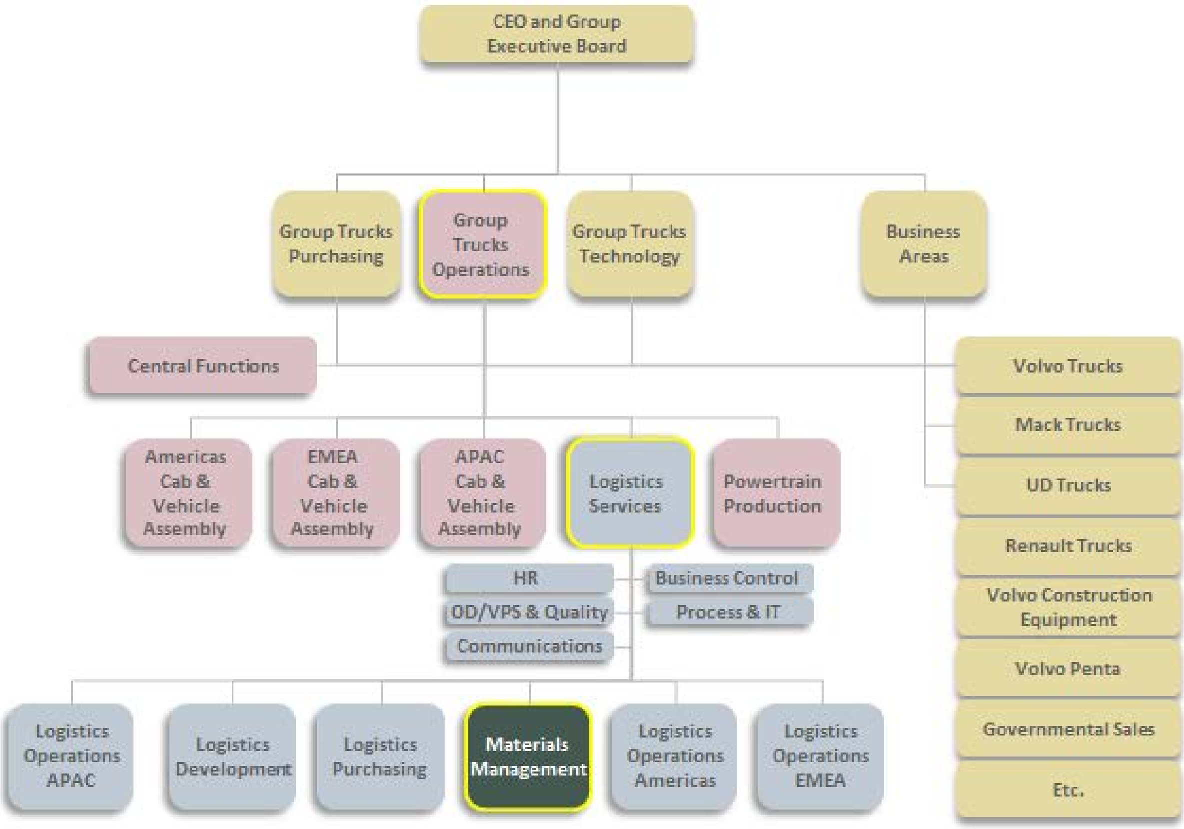

FIGURE 6: A TOP-VIEW ORGANIZATIONAL STRUCTURE OF VG. (VOLVO GROUP, 2016B)

Group Truck Technology (GTT) are solely responsible for research and product development of complete vehicles, powertrain, components and service offering.

Group Truck Purchasing (GTP) is the global group function covering the purchase of automotive products and parts including aftermarket, for all truck brands in VG.

Business Areas include all the brands and products offered by the entire VG organization, which can be viewed in figure 3. It also includes services such as such as governmental sales and financial services. Each area is responsible for their own sales function as of Mars 2016.

Group Truck Operations (GTO) are responsible for manufacturing of cabs and trucks for the Volvo, Renault Trucks, Mack and UD Trucks brands as well as production of the VGs engines and transmissions. Additionally, they are responsible for spare part supply to the VGs customers and logistics operations. Which falls upon the subsidiary body called Logistics Services (LS), where their aim is to establish and secure "... global availability of aftermarket parts to dealers and end customers at the right time, the right place and at the right cost" (Volvogroup.com, 2016 - A). Thus LS is responsible for the management of spare parts in VGs distribution centers around the world, both operationally and managerially (Volvogroup.com, 2016 - A). This includes the design, management and development of the entire supply chain network. Thus LS main responsibilities are predetermining stock and service levels and also forecasting and optimizing said inventory levels throughout the supply chain. Within LS the management of VGs aftermarket, for all brands, is directed at the Material Management (MM) division: who are responsible for the material flow in aftermarket supply chain, from suppliers to end-customers.

4.2 VOLVO'S AFTERMARKET SUPPLY CHAIN

The MM division, a part of GTO, are responsible for the aftermarket supply chain, is process oriented where the flow both physical and informational follows certain work processes. The different processes are part of employee's operational tasks and the VGs company culture encourages the improvements of said processes, in order to add additional value to their working methods. The MM divisions' organizational structure can be seen in figure 5.

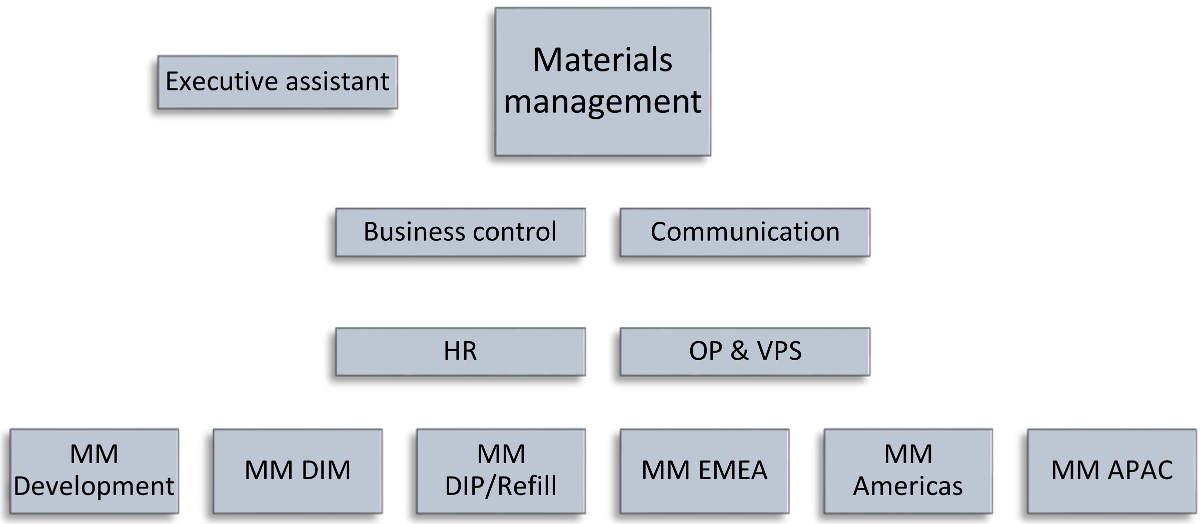

FIGURE 7: ORGANIZATIONAL STRUCTURE OF THE MATERIAL MANAGEMENT DIVISION. (VOLVO GROUP, 2016B)

VGs aftermarket supply chain, which can be viewed in figure 6, uses a decentralized structure with a complex network of nodes. Approximately 5000 suppliers ship spare parts to 6 central distribution centers (CDC) around the world, in turn the CDCs supply parts to 10 support distribution centers (SDC) and 80 regional distribution centers (RDC). All of which serve a network of roughly 3000 dealers, authorized by VG. The physical flow of the supply chain mainly goes downstream to the end-customer through the dealers, however there also exists a return flow upstream from dealer to the different distribution centers (CDC, RDC and SDC). Additionally, there is an information flow which moves upstream, the information is related to the ordering of parts.

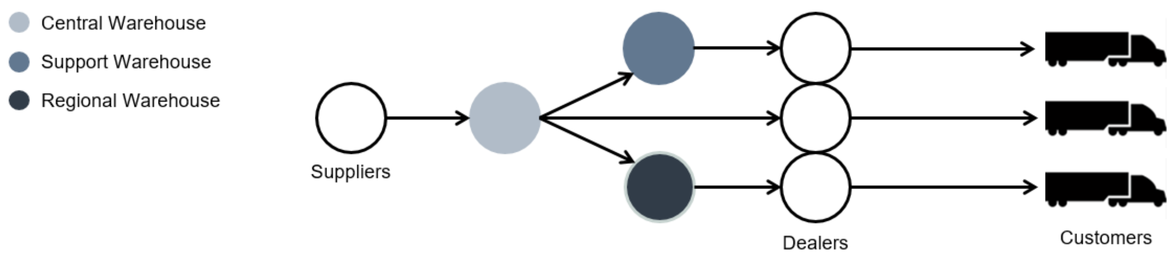

FIGURE 8: VG'S AFTERMARKET SUPPLY CHAIN. (VOLVO GROUP, 2016C)

The six CDCs are currently located in Europe, the Americas and Asia. The largest of which is located in Gent, Belgium. It contains approximately 200,000 SKUs and handles about 6-7 million order lines per year, put in the reference of the entire aftermarket supply chain which handles roughly 26 million order lines per year. The CDCs receive spare parts from VGs supplier network and their purpose is to supply parts to the two different distribution centers, RDC and SDC, and directly deliver to dealers. The RDC are setup as smaller versions of the CDCs with the aim to maintain inventory for a specific geographic area which would have a too long lead time if distribution was done through the CDCs. Its purpose is to supply dealer with necessary inventory in order to minimize lead time to end-customers when there is no available stock at the dealers. There are no RDCs in the Nordic and European region because of the CDC located in Gent. In Europe and the Nordic markets there are instead a network of SDCs which stocks inventory with a high demand frequency in order to supply dealers with day-to-day orders, basically distributing emergency order to dealers.

The MM division is currently focusing and working towards the development of current planning processes. Future concepts which are being discussed within the division are using POS data from dealers as a basis for forecasting stock levels throughout the aftermarket supply chain. This would, according to VG, reduce lead times and internal costs by reducing safety stock levels and minimize occurrence of over stocking. A necessity would be integrated planning for the different nodes in the distribution network which aims at to create a more agile forecasting.

Within the MM division there is a monitoring and control of certain KPI which can be viewed in table 9. Each KPI can be related to several planning processes which stretch over the entire aftermarket supply chain.

TABLE 9: SHOWING ALL KPIS AND PI WHICH IS PART OF MM'S BUSINESS FUNCTION. GREEN BOXES ARE DIRECTLY LINKED TO KPI OF LS AND YELLOW BOXES ARE INDIRECTLY LINKED. (VOLVO GROUP, 2016B)

Category	*Name of KPI*
Quality	ETA Quality
Quality	Forecast Quality
Delivery	Aftermarket Parts Backorder recovery
Delivery	Aftermarket Parts Availability
Delivery	Dealer Service index
Delivery	Dealer Inventory Mgmt coverage
Delivery	Aftermarket supplier delivery precision
Cost	Freight cost per volume
Cost	Aftermarket Inventory days
Environment	Air share for refill
People	Improvements per employee

4.3 THE AFTERMARKET SUPPLY CHAIN PLANNING PROCESS

The planning process for spare parts control within MM can be divided into four main roles:

- Continental Material Planning (CMP) who manages the replenishment of parts to the CDCs.
- Demand and Inventory Planning (DIP) manages the material flow in CDC by setting forecasts.
- Refill Material Management (Refill) who manages the spare parts flow from CDCs to RDC and SDC.

- Dealer Inventory Management (DIM) that works with dealers in managing their inventory of spare parts, both independent and directly dependent dealers.

A simplified schematic flow can be viewed in figure 8, which shows how the demand in the aftermarket supply chain is flowing.

FIGURE 9: SHOWING THE FLOW OF DEMAND IN VGS AFTERMARKET SUPPLY CHAIN. (VOLVO GROUP, 2016D)

4.3.1 DEALER INVENTORY MANAGEMENT

The purpose of DIMs planning processes is to secure availability of spare parts on a local level for individual dealers. Thus supporting and securing end-customer satisfaction, consequently, uptime. It is the only part of the whole planning process which measures and follows actual demand from end-customers. Because of this it is the only process which has relatively accurate sales data at dealer, in other words POS data. Its aim is to cost efficiently supply spare parts to dealers and at the same time keep a high service level, thus balancing distribution and capital costs. With each dealer in VG distribution network a Logistics Partnership Agreement (LPA) is signed which sets the terms of each party's responsibility, in some cases VG has full control over the inventory at dealers and in other cases the agreement is purely transactional. Furthermore, VG conducts regular buybacks with dealer, meaning that obsolete stock is repurchased by VG. DIM has 4 key performance indicators, all part of the MM divisions PI. They monitor service index, turnover rate, healthy stock and automatic share of stock at each dealer. Additionally, DIM secure performance of dealers by optimizing inventory parameters, such as safety stock (SS), economic order quantity (EOQ) and reorder point (ROP), and by extensively coaching dealer in their respective market.

The current planning process in DIM is based on conventional forecasting methods on historic sales data received from dealers. Each month new forecasts are made on a part number level; this includes recalculating the levels of SS, EOQ and ROP. The supply process at dealers is done in two ways, either through orders executed by the dealers themselves when there is a need or the supply is controlled by VG. The former type of order usually have one day of lead-time. The later type orders new stock through an automated process based on the ROP. For dealers there is no monetary distinction in the two ways however for VG the former is more costly because spare parts has to be delivered, usually urgently, from the closest distribution center (either RDC, SDC or CDC).

At VG a system is used for coordinating and managing inventory at dealers. This provides responsible employees at VG a tool for communicating and maintaining availability at dealers, which is done by DIM. The system which is updated by dealer usage sends data regarding sales, order picks and inventory levels on a daily basis. Additionally, every weekend data is aggregated for every article number which is sold through that particular dealer.

4.3.2 REFILL MATERIAL MANAGEMENT

There are two processes within VGs internal aftermarket supply chain that handle the spare parts flow. One of which is Refill and the other DIP, the latter will be extensively presented in chapter 4.3.3. Refill manages the material flow between the CDCs and RDCs or SDCs. The main focus of this planning process is to reduce supply chain costs, increase availability to dealers and balance the tied up capital in regard to order handling costs. These goals are based on three principle policies:

1. The stock holding policy, i.e. what to stock and what not to stock. Which is based on how many order picks has been conducted in the last 12 months and what the stock value per piece is. Employees at refill uses a matrix in order to decide whether to stock according to the EOQ, only stock one piece or adjust the current EOQ.

2. The refill policy, i.e. when to order and how much. This policy is entirely based on the ROP and EOQ, where the ROP is calculated on the basis of the decided SS levels and lead-time consumption. The amount which is to be stocked is decided according to the EOQ in addition to stocking up to the ROP.

3. The return/scrap policy, i.e. when to return and when to scrap. It is part of Refills role to manage the reverse logistics of spare parts. Which in turn is done at RDC or SDC levels, each distribution center has its own return policy.

Additionally, a significant task of Refill is to secure the system lead time of spare parts, this is done by monitoring deviations in lead time, optimizing, validating and setting correct lead times in the systems and also following up on actual lead time discrepancies. Lead time at Refill is viewed as an important parameter in the spare parts flow because the longer or the more unstable lead-time the higher risk there is of negative impact on the availability, inventory and air freight costs.

4.3.3 DEMAND AND INVENTORY PLANNING

The activity of DIP process for the CDC, in Ghent, is performed by the DIP team which is in the scope of this study. DIP for Ghent distribution center operates from three locations: Gothenburg, Eskilstuna and Lyon where each location is responsible for certain business areas. Gothenburg's team is responsible for demand and inventory planning of Volvo Trucks (VT) and Volvo Buses as well as Volvo Penta, where VT is the scope for this study.

Planning process
The DIP team at VG is in responsible of the demand and inventory of the CDC at Ghent. The regular tasks for the DIP team are related to:

1. Forecasting spare parts.
2. Lead and contribute to the flow planning work.
3. Identify, analyze and rectify according to seasonal sales behavior.
4. Perform data analysis and utilize it to identify root causes.
5. Supersession.
6. Phase in coordination from projects, including set initial stocking and registering the new part.
7. Scrapping obsolete parts.
8. Provide estimation for “All Time Buy” quantity when supplier can't provide the item.
9. Regular meeting with different stakeholders in VG responsible for other business areas.

Classification
The different roles at the DIP team are related to the classification of spare parts made by VG. The classification is based upon the parts price, the frequency of the orders and its criticality. The DIP team also considers what stage of the life cycle the item is.

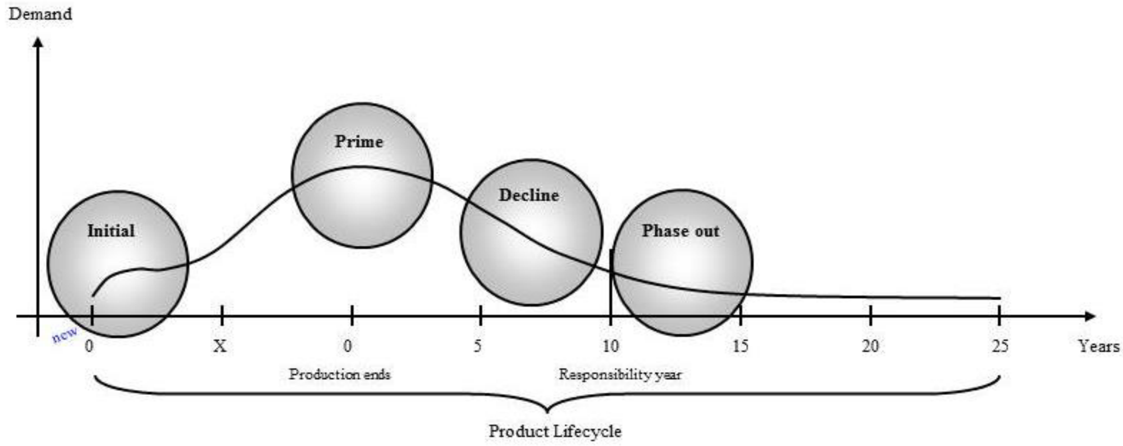

FIGURE 10: SHOWING THE LIFE CYCLE OF SPARE PARTS, AS VIEWED BY VG. (VOLVO GROUP, 2016A)

From figure 9, the initial phase is when a completely new part is entered into the aftermarket supply chain. In this phase the DIP team is deciding whether the certain spare part will have an initial stock level or not and the quantity. The prime phase is when the part has a steady demand, it is then critical to ensure high availability downstream. The decline phase represents the phase when the demand is declining and production of new vehicle with the item has ended. In this phase the inventory levels will be reduced in order to minimize the risk for the part to become obsolescence, without affecting the service level significantly. The last stage, in figure 9, is the phase out and is when the part has exceeded the responsible year which is the time that VG guarantee the customer that the supply of spare part will be available. This mean that the focus is to lower the inventory and the service level for these and if it's not considered profitable to still provide the service level will cease to exist, meaning no supply and inventory. Table 10 is a summarization of the categorizing of spare parts for both the frequency, phase and price. The price is represented by the second letter in the designation where "A" is the lowest price segment and the highest letter has the highest price, i.e. X.

TABLE 10: SPARE PARTS CLASSIFICATION, A IS THE LOWEST PRICE SEGMENT.

Segments	Initial phase	Prime phase	Decline phase	Phase out
Initial parts	0C			
Critical parts	AA-AI BA-BI CA-CI DA-DI	AA-AI BA-BI CA-CI DA-DI	AJ-AK BJ-BK CK-CK DJ-DK	AL BL CL DL
Fast Moving		EA-EE	EA-EE	EA-EE
Fast-Medium Moving		EF-EJ	EF-EJ	EF-EJ
Medium Moving		EK-EO	FA-FC FM	GA GX
Medium-Slow Moving		EP-ET	FD-FF FN	GB GX
Slow Moving		EU-EZ	FG-FI FO	GC GX

Forecasting process
The forecasting period at VG is split into 53 periods in one year and the horizontal stretches over a year with the most important within three month, which is set in regard to the average supplier lead time. The aftermarket at VG contains huge number of SKU and the planning of material flow in VG is mainly determined automatically by IT tools. This includes setting an appropriate inventory level, order quantity, forecasting and scheduling. The system automatically interrupt the decision and notifies the user for manual intervention, which is the main activity related to forecasting for the DIP team, where they investigate any deviation notified by the system. These deviations are notified when an exceptionally large change in the demand for a spare part number has occurred in short amount of time.

The forecast is based on historical demand and are created by the use of double exponential smoothing, which considers the trend of demand. Other forecasting methods are used for lumpy items with very low demand and for newly introduced items. When back-testing the exponential smoothing for one of the fast spare part it was shown to be delaying when the demand is ramping up and the lagging is seen over the whole phase. A graph is shown in figure 10.

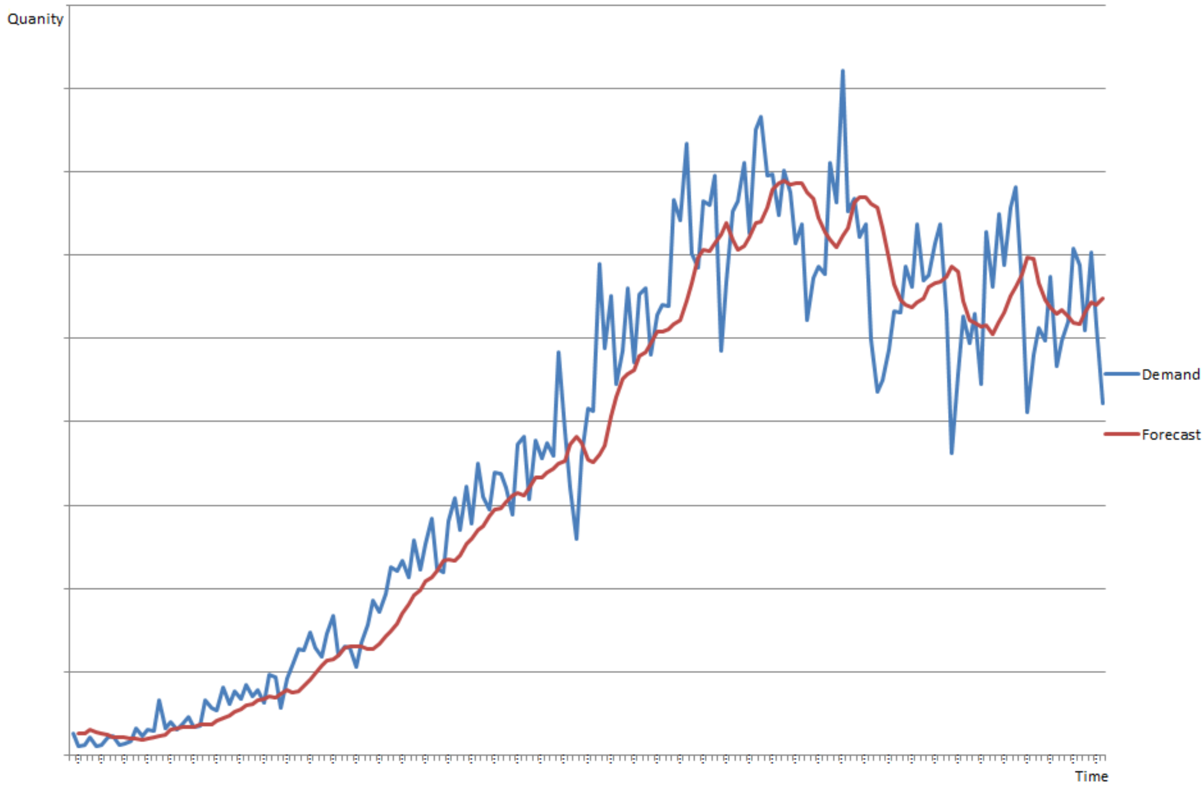

FIGURE 11: SHOWING A BIAS FORECAST OF A SPARE PART THAT HAS SEEN INCREASING DEMAND OVER SEVERAL YEARS.

The DIP team claims that they have problems with the forecast accuracy when demand is suddenly increasing over a long period of time. They have a vision of being more proactive rather than reactive. Currently there is a limited view on the actual end-customer demand, the only information received is from sales department, with data regarding total expected sales for the whole aftermarket for a certain vehicle model, which isn't useful for potential improvements in the forecast since the data is not detailed. Some information regarding campaigns is shared regarding specific markets but is also not used commonly since the change in demand for specific markets hardly change the total demand of an item. The forecasting is based on historical data of demand and no information regarding the POS data from DIM is shared within the supply chain. In figure 11 a comparison of demand for a spare parts number between DIM and DIP is shown. Which indicates a bullwhip effect in the aftermarket supply chain.

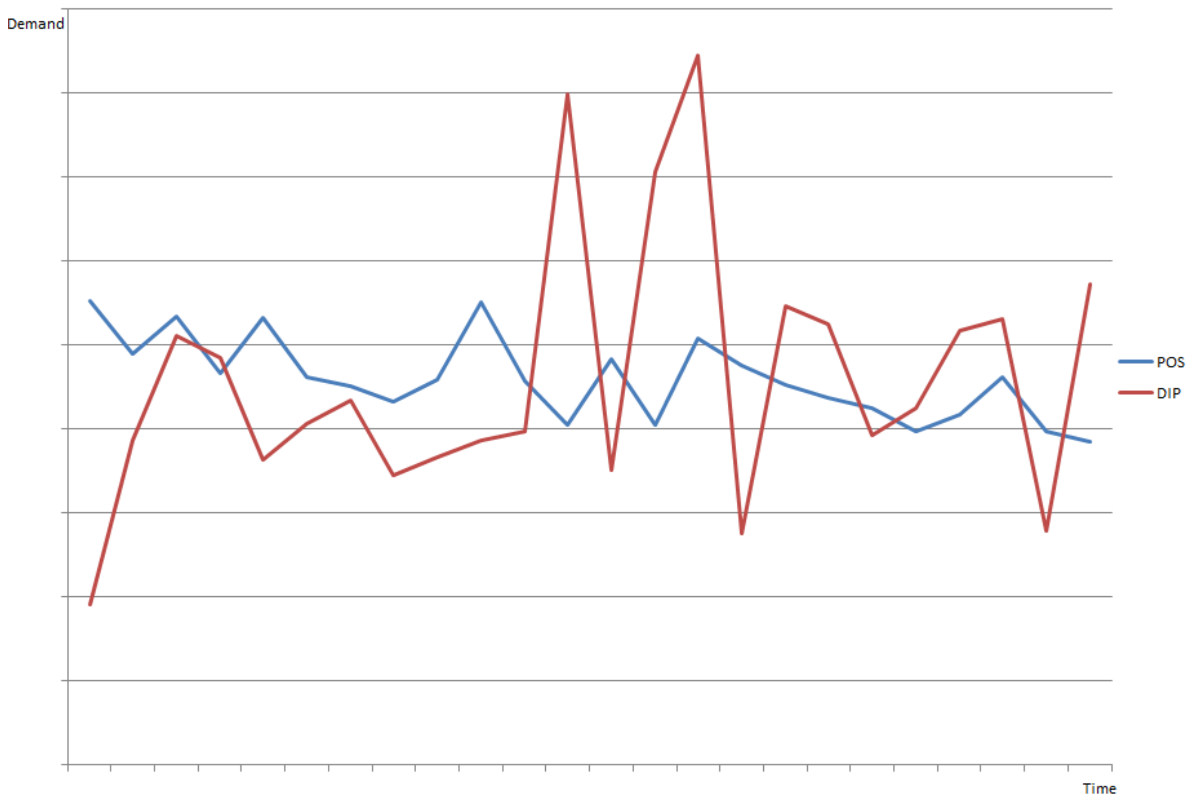

FIGURE 12: A SPECIFIC SPARE PARTS DEMAND PATTERN IN DIFFERENT STAGES OF THE SUPPLY CHAIN.

Key performance indicators

The DIP team's KPI is forecasting quality, also called forecast accuracy. The comparison is made between the initial forecast made the first day of a period for the next period in relation to the actual outcome. The team compares the forecast quality for the forecast made in the end of the period with the one made in the beginning of the period. This is conducted in order to see what the manual activities have contributed with and how much it is improved by the DIP team. Other KPIs of interest for the DIP team are the inventory value, inventory turnover and availability. This is measured in terms of availability for the demand received by the refill team and dealers.

4.4 BIG DATA AND CONNECTIVITY AT VG

It has been identified at VG that connected vehicles through telematics is a future area which is top prioritized by the company. Furthermore, based on previous benchmarking at VG with competitors, the usage of big data and analytics tools is a top priority for all vehicle original equipment manufacturers (OEM). Since 2015 the automotive industry is the 2nd largest generator of data, according to VG. Currently, VG is realizing the opportunities which arise from connected vehicles; based on telematics installed with more than 70 sensors. VG has almost 700,000 connected vehicles and that population is increasing with approximately 15,000 vehicles per month, with the goal to have a completely connected fleet. The purpose of such an IT solution and utilization of the gathered data is to support and help facilitate the vehicle uptime of end-customers. It has been found that VG is in a similar stage as most competitors (65 %) when it comes to maturity of utilizing big data. There are lots of decentralized activities with pockets of excellence which use available technology and data on an Ad-Hoc basis, however not on a cross-functional level. Data retrieved from telematics and the aftermarket supply chain is not governed on a centralized level.

Moreover, there is a lack of business models with customer focus for analyzing big data and creating partnering or engagement models. Furthermore, data is not stored with a common platform and there are data quality issues. The fleet of vehicles connected to VG through telematics gives great amount of real-time data input from end-customers; however investigating opportunities for said data which improve the current planning processes for spare parts at LS is currently in its initial phase. This development is quite interesting in order to increase efficiency in the internal supply chain and the potential positive impact of such an implementation is recognized as a driver for changing the current planning environment for the aftermarket on a global scale.

4.5 AVAILABLE DATA AT VG

An extensive study of VGs systems and databases was conducted in order to map data flow and also identify any available data which could positively impact forecast accuracy. From the system mapping it was identified that five sources could be of interest. Each data source deemed relevant for potentially improving forecasts will be thoroughly presented. The data sources are all part of the vehicles life-cycle, which are shown in figure 13.

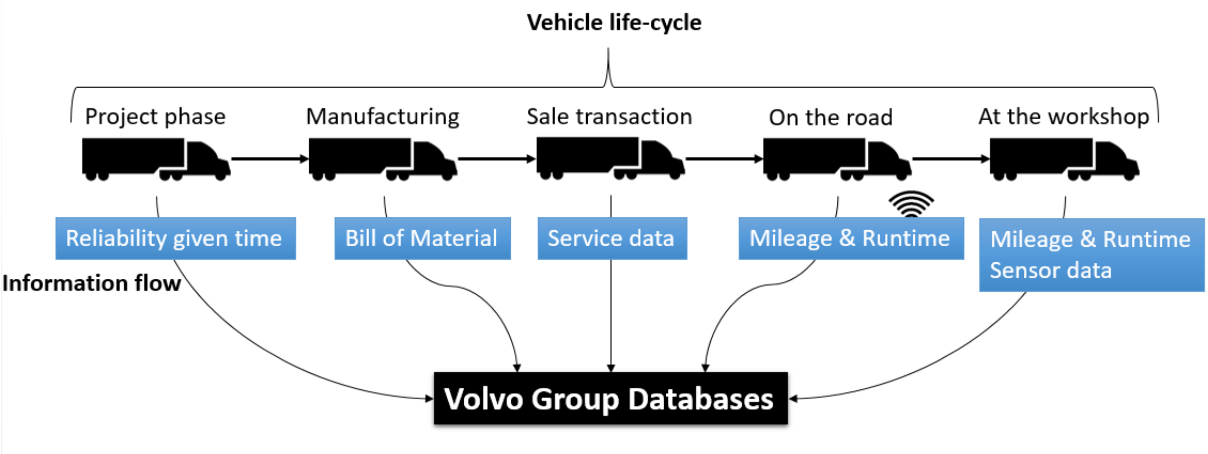

FIGURE 13: STAGES IN THE VEHICLE LIFE-CYCLE, ALL OF WHICH ARE SOURCES OF DATA.

4.5.1 PROJECT PHASE

GTT is conducting experiments on components and systems which exist in trucks in order to estimate the lifespan of spare parts, called reliability given time (see figure 13). This is done through a series of pressure and stress test, by for example shaking components or systems until breakdown. Each test is based on estimating the mileage or runtime, however it is unknown in what extent the amount of different spare parts are tested. Some critique is raised to the estimates' reliability as the parts are "built-for-life". It is becoming more common that such tests are the domain of suppliers who deliver the parts to VG. Additionally, physical tests are decreasing with an increased focus on computer simulations.

4.5.2 MANUFACTURING

By generalizing the information flow of data from the manufacturing phase it has been found that at VG there are two aspects to consider. Either data in reference to a spare part number (identification of a spare part in the aftermarket) or a production part number (identification of a spare part in production).

- Spare part numbers is stored for the purpose of providing the aftermarket with a common reference and point of view. It is mainly used in a global spare parts planning system. This also contains data regarding cross referencing between spare part numbers and production part numbers.

- Production part numbers is used for providing the engineering departments with a common reference point. The product data management system (PDM) which is used specifies the entire bill of material (BOM) for different truck models. In addition to the PDM there is data storage of a production part number for specific truck configurations. This in effect is dependent on the truck's engine and working environment.

Furthermore, there is manufacturing data in reference to vehicles, which uses a unique identification number for individual vehicles, called chassis number. Data which is connected based on chassis number includes weight of truck, brand type, components and variant of truck.

4.5.3 SALES TRANSACTION

There is also data storage which deriving from the sales transaction of a vehicle. This includes customer oriented data, such as:

- Buying and using customer contact information.
- Operational information such as country of operation and associated dealer.
- Sales person of vehicle.
- Delivery information, such as date, dealer and importer.

Additionally, certain spare parts are referred to as service spare parts. Which is included in the trucks service agreement established in connection with the sales transaction. Currently VG has 3 different levels of service agreements: gold, silver and bronze. The first of which is a premium service agreement which offers 100 % uptime for customers where VG is responsible for any costs related to vehicle downtime. However, it was pointed out that there is a distinction between what is pre-planned for exchange and what is included in the service agreement. Based on the conducted studies at VG only spare parts classified as filter and oils are pre-planned. Meaning that these are the only types where there is an estimated date of service. This is planned on a time horizon reaching for about two years, however it was concluded that such dates could differ significantly depending on if the customer choose to follow the established service plan. The service agreement only specifies who is responsible for the cost of a spare part, thus if it is included then VG is responsible for all cost.

Currently at VG there are two systems which is used for using stored data regarding service for vehicles. One system is used for planning service work for individual trucks based on previously mentioned chassis numbers, as mentioned earlier only oils and filter. This uses data regarding the truck, such as engine size and working environment, in order to correctly plan the intervals for service. Such a service plan is initially set by sales department in accordance with the customer needs, however it was identified that such initial plans are changed and optimized with the aid of better knowledge from employees only working with service planning. Service intervals are measured in mileage or runtime of the truck. Another system is used for logging any service conducted by workshops on end-customer vehicles.

4.5.4 ON THE ROAD

Wireless transfer from connected vehicles is done through telematics technology which is installed in all newly produced vehicles. The wireless transfer began 2007 for customer that wanted to have connected vehicle to monitor. Since 2014 all new produced vehicles transfer vehicle and are currently approximately 700,000 vehicles has wireless transfer capability. The data sent from these vehicles is considered as main data which is limited to fuel consumption, mileage, run time and error codes sent several times per month, the data sent is limited because of the costs associated with the telematics technology. This uses cellular connection for transfer and is considered as a high cost. Thus only vehicles owners who have paid for extended services for monitoring send data daily and the data is logged inside the vehicle every five minutes which then is uploaded to VG daily. Data included in this extended services is limited to 10-15 data parameters, which includes the mentioned main data. Currently this is limited to 11,000 vehicles in the VT brand. In addition to main data VG also receive data regarding battery state of health, brake pads status, clutch usage and air cartridge status from these 11,000 vehicles. This process also contains data regarding the positioning of said vehicles. Furthermore, it was stated that additional vehicle data from the connected vehicles can be extracted upon requests through wireless transfer but which data that can be extracted is limited to how the ECU software is programmed to store data, this type of extraction is available for VGs internal stakeholders and to change what data is to be stored requires certain business processes on higher decision levels which is reviewed approximately twice per year. It was concluded that the data quality of all connected vehicles through wireless transfer is considered to be high in relation to main data (fuel consumption, mileage, error codes, and runtime), other data such as sensor data was considered to be of lesser quality.

The most significant purpose for usage of the vehicle data that is received by the 700 000 vehicles is to identify quality issues in vehicles, either by individual components or entire systems, which in turn is dependent on the vehicles working environment. Furthermore, projects regarding vehicle fuel consumption and driver behavior in order to maximize fuel efficiency uses the presented vehicle data as basis for analysis. The additional vehicle data that can be requested for the purpose of specific business cases are most related to a proactive approach for vehicle uptime. Such business cases are first and foremost related to customers who have purchased extended services for their vehicles. From VG it was identified that there is developing projects which aims to consolidate vehicle data in order to predict occurrence of breakdowns. Combining said data with tools used for statistical analysis gives, as stated by VG, from 80 % certainty that breakdowns will occur within the coming month. Consequently the purpose is to support the uptime of vehicles.

4.5.5 AT THE WORKSHOP

Workshops manually extract data from the vehicle by physically connecting to the vehicle's ECU and uploading this into VG servers. This is done every time a vehicle visits a workshop. This data has the highest quality and involves most data parameters which VG has logged data since 1999. This process can take a couple of minutes or several hours depending on how long the vehicle is connected to the extraction tool. Each data point which is extracted has a set priority that specifies to what extent the data is extracted depending on the time of transfer. It was stated that data points with higher priority is more reliable because of its continuity of extraction. Data which is transferred with this process includes mileage and runtime readings, fuel consumption, error codes of components and systems and vehicle sensor data (70 sensor points), where mileage, runtime and fuel consumption has the highest priority. This type of data transfer is considered low cost because it uses internet as means of communication, which provides larger bandwidth and consequently larger datasets can be transferred.

4.5.6 NOT INCLUDED IN VEHICLE LIFE-CYCLE

An additional data parameter not considered part of the vehicle life-cycle was found at VG. It was identified that customer loyalty has a direct impact on the demand for spare parts. Loyalty of customers indicates as to what extent it is preferred to choose the available Volvo branded spare parts instead of a competitor's. However, this is measured in terms of end-customer satisfaction in reference to products and services offered by VG through surveys. Currently at VG there exist two such surveys, one conducted by VG for internal improvements towards dealers and workshops. This survey is done manually by VG where personal interviews are conducted by phone with approximately 90,000 end-customers. One aspect of the survey is measuring the customer satisfaction of availability of spare parts. This is done in relation to their respective dealers/workshops, however it was stated that such data is difficult to translate because of its subjective nature. In addition to this a survey conducted by an external company, in cooperation with all other competing OEMs. It is used as a benchmarking for all OEMs where the end results shows which aspects of services offered are important for the customer, in terms of drivers for change. It also shows where the competing brands stand in relation to each other concerning customer satisfaction. This survey also measures opinions of participants in regard to availability of spare parts; however it is conducted on a smaller scale (790 interviews with end-customers).

5. RESULTS AND ANALYSIS

Initially, this section gives a description of big data's potential in improving demand planning for spare parts in the aftermarket. Furthermore, two dimensions which effects the planning process of forecasts are described. Finally, a framework describing different levels of maturity in adopting big data is presented.

5.1 POTENTIAL FOR BIG DATA IN FORECASTING

In the aftermarket supply chain of an automotive company, such as at VG, there is a general view of endless potential for improving forecasts through the use of big data. Currently their forecasts are entirely based on historical sales data. Thus any external data which can help facilitate the planning processes for the aftermarket can be considered to have a significant improvement in forecast accuracy. This is also backed by academia which explains that forecasts should not solely be based on historic demand but also use data input from other sources in the focal company as it has the potential for greatly improving results through capital gains and customer satisfaction (Kennedy et al., 2002; Baudin, 2004; Ton, 2013; Hassani and Silvia, 2015; Accenture, 2014). Subsequently at VG there is no exploitation of data created in the aftermarket supply chain for improving forecasts. Considering that Chopra and Meindl (2013) describes that an aftermarket supply chain responds based on a demand, usually as soon as possible. Therefore, an aftermarket supply chain can be viewed as an agile supply chain because it needs to handle quick responses and adapt to customer demand. Having supply chain visibility downstream is of the utmost importance when having an agile supply chain (McIntire, 2014). It provides the focal company a way to make sure that spare parts is in the right place at the right time. Creating a match in demand and supply and thus mitigating the incongruity which is created by the large amount of SKUs in an aftermarket supply chain. Considering that spare parts can be viewed as one of the most profitable sources of income for manufacturing companies (Wagner and Lindemann, 2008), it is easy to understand that accurate forecasts for directly affects a company's profitability. Either by minimizing necessary inventory or through minimizing loss of sales at customers.

As observed at VG case there exists a wide gap between knowing what data is available and where that data is located. Considering the size of automotive companies such as VG there are large amounts of knowledge in regard to technical and business know-how. Which if given the opportunity can exploit big data for improvements in areas such forecasting. However, the sheer size creates silos with difficulties in creating cross-functional communication and collaboration. This can be an issue in any automotive company which wants to adapt a more visible supply chain, as stated by research that studied 150 companies who despite being in the top of their field need to get rid of silos and create a more holistic view (Cavalieri et al., 2008; Aberdeen Group, 2006). To successfully implement the usage of big data in planning processes for the aftermarket a disruption is needed in current work processes for parallel business functions who owns the needed data.

The results from the case study showed that big data offers great potential in order to improve forecasts for the demand of spare parts. For example, the empirical study describes available data at VG and also current views and opinions of the potential for big data. Combining different data sets gives insight into what the actual demand is. This is a way to short-circuit the information flow in the aftermarket supply chain which supports an increase in supply chain visibility. Using historical sales data as a basis for forecasting in a supply chain which has multiple levels of supply and demand creates significant bullwhip effect which can be viewed in figure 8. By short-circuiting the information flow such bullwhip effects can be mitigated. In order to be successful with such an implementation, large amounts of data is needed from and in regard to the end-customers. Large datasets provides a more accurate picture of true demand (Mayer-Schönberger and Cukier, 2014). Furthermore, datasets which are received from connected vehicles, through telematics for example, creates opportunities to conduct preventive maintenance and real time diagnostics. This supports the supply chain visibility of the aftermarket through increasing customer satisfaction and agility, ultimately increasing the uptime of vehicles.

5.2 CAPTURED DATA

In the empirical study, chapter 4.5 presented available data at VG which had a potential for improving forecasting. According to McIntire (2014) the first step in achieving supply chain visibility is capturing data, called sensitivity. However, potential improvements derived from the captured data depends on the quality of capturing process. In order to successfully translate big data into a demand, in accordance with equation 4 from chapter 3.6 in the theoretical framework, the quality of the capturing process need to be considered. For conveniences purposes equation 4 is presented again below. Each variable in equation 4 has one or a combination of several corresponding data parameters found in the empirical study.

$$D(t) = K\,(P_1(x_1(t)) + \cdots + P_n(x_n(t))) + \varepsilon$$

EQUATION 4: TOTAL END CUSTOMER DEMAND BASED ON LIFE DATA ANALYSIS.

The empirical study showed that no available data at VG is shared to or used in the planning process of spare parts. Meaning there is no accessibility, intelligence or decision-relevance for any of the data parameters which are interpreted as variables in equation 4. This implies that the each scorecard table presented in this chapter has a rating of zero in each of the three succeeding steps for achieving supply chain visibility.

5.2.1 POPULATION AND LIFE DATA

Population, variable *n* in equation 4, refers to all spare parts of a specific spare part number that are currently in use. For the entire population there is need to determine life data, denoted as the function *x(t)* in equation 4, which is any data related to each specific spare part that has a correlation with occurrence of failure. Life data could be considered as the most fundamental data parameter that a company needs in order to make accurate forecast of demand based on big data, if using equation 4. Population has one interpretation as defined earlier in this section and life data has been described by Nelson (2005) as anything related to the usage of a product and its correlation to a failure, for example hours, month, cycles etc. When understanding the total demand it is therefore preferable that life data exists for the entire population. It is easy to comprehend that having several hundred thousand SKUs and a population of vehicles which stretches up to half a million for just one SKU creates massive amounts of data, big data, which must be structured in order to create intelligence and facilitate decision making.

Time of usage
The amount of vehicles on the road and date of registration combined with bill of material gives the population of a spare part number, but also how long it has been in usage i.e. life data. This is captured within VG with some lack in quality, since the time a spare part has been in usage is not reset whenever it is replaced with a new one. This motivates the sensitivity score viewed in table 11. However, there are still possibilities to reset the time for replacements by using the historical sales data as a reference point for when new parts are sold.

TABLE 11: THE CURRENT SUPPLY CHAIN VISIBILITY SCORECARD FOR SPARE PARTS TIME OF USAGE AT VG.

Parameter	Sensitivity	Accessibility	Intelligence	Decision Relevance	Fitness
Time of usage	3/4	0/6	0/4	0/10	13 %

Mileage and runtime
The received data parameters, mileage and runtime, could be used as a way to express a population and its growth in time. Consequently runtime and mileage can be viewed as life data, because those data parameters affect the breakdown frequency. Mileage or runtime in essence describes the usage of a vehicle and thus the usage of a spare part. Data from vehicles that doesn't have wireless connection can be considered to have lower quality since the data is less frequent and it doesn't include vehicles that didn't conduct maintenance at a VG workshop. Additionally, there is an opportunity to use linear extrapolation to determine the data values that are in between the measured values. This could be used when back-testing different theories on old data to validate the estimated demand data given. When estimating future demand the mileage and runtime could be estimated by linear equation. For example, (1) if the vehicle has driven on average 1000 km per month historically and the last received measured mileage is 2 months in the past, 2000 km could be added on the last known mileage. (2) If the wanted estimated demand is two month in the future the vehicle has driven 1000 km per month historically, 2000 km could be added to current mileage.

Significant knowledge and data regarding mileage and runtime of vehicles was found at both the sales department and GroupIT. The data is complete and frequently updated, several times a month for connected vehicles, but with some quality issues that could be considered predictable. For the case of VG data from wireless transfer of connected vehicles covers a portion of the population, but there is still an unknown portion of the population which is not connected. This part of the population can however be assumed to be part of the manual transfer of data conducted at workshops. Wireless data extraction alone would give the highest sensitivity score and manual extraction through workshops the second lowest score, combining these gives the sensitivity score presented in table 12.

TABLE 12: THE CURRENT SUPPLY CHAIN VISIBILITY SCORECARD FOR MILEAGE AND RUNTIME AT VG.

Parameter	Sensitivity	Accessibility	Intelligence	Decision Relevance	Fitness
Mileage and runtime data	3/4	0/6	0/4	0/10	13 %

Sensor data
Another data parameter found at VG that could be viewed as life data, is sensor data from the vehicles. The analysis of the empirical data show that the correlation between breakdown and sensor data is something that VG has put significant focus on, in order to predict occurrence of failure to achieve preventive maintenance rather than corrective, ultimately to secure uptime. However, sensor data from the vehicles is limited to manual extraction of data and when making specific request of data parameters through wireless connection, even in these case, the data is normally only received on limited population. There are currently just a few spare parts that have shown any correlation between sensor data and failure, the correlations are often complicated and require significant research often based on error codes in combination with sensor data. The current state of sensor data is that it’s captured but incomplete, thus a sensitivity score is presented in table 13.

TABLE 13: THE CURRENT SUPPLY CHAIN VISIBILITY SCORECARD FOR SENSOR DATA AT VG.

Parameter	Sensitivity	Accessibility	Intelligence	Decision Relevance	Fitness
Sensor data	1/4	0/6	0/4	0/10	4 %

5.2.2 PROBABILITY DATA

The probability *P*, in equation 4, referees to the likeliness that a spare part breaks, the preciseness of this variable is important in to achieve high accuracy of the estimated demand. The probability of failure is determined by life data. Probability could be determined either when the data is "complete" meaning that the occurrence of demand is known and the probability of failure is either zero or one. Else, the probability is between zero and one and can be determined based on distribution of probabilities, such as Weibull and Gaussian distribution. VG are currently working with determining the probability of failure in a way which can be called "reliability given time". It refers to the percentage that a spare part will break down before a given time. Example; there is a 90 % probability that a spare part will operate successfully after 1000 miles. Additionally, there are two more ways to determine the probability of failure, based on service data or technical life length.

Service schedule
The schedule is stored at VG and continuously updated based on real time data from vehicles, but only covers the population that bought additional services, which represents 11 000 vehicles. The data itself could be seen as "complete" data since the probability *P* is either one or zero, but since the data only covers small a portion of the population, for the purpose of improving forecast it could be seen as stored but incomplete. This gives a sensitivity score as viewed in table 14.

TABLE 14: THE CURRENT SUPPLY CHAIN VISIBILITY SCORECARD FOR SERVICE SCHEDULE AT VG.

Parameter	Sensitivity	Accessibility	Intelligence	Decision Relevance	Fitness
Service schedule	1/4	0/6	0/4	0/10	4 %

Service interval
VG has service intervals for the rest of the population, beyond the 11 000 vehicles. Intervals are only based on the vehicle configuration and are not adjusted based on any additional data. This mean that this data parameter has low quality and can also be viewed as stored but incomplete, giving the same score as service schedule which can be seen in table 15.

TABLE 15: THE CURRENT SUPPLY CHAIN VISIBILITY SCORECARD FOR SERVICE INTERVAL AT VG.

Parameter	Sensitivity	Accessibility	Intelligence	Decision Relevance	Fitness
Service interval	1/4	0/6	0/4	0/10	4 %

Technical life length data
Other data related to the probability of failure is the "reliability given time" of a spare part. The reliability given time is determined through test made by VG in order to estimate a technical life length. This could be estimated for a component or for systems, such as a whole motor, shaft, dashboard etc. The data were considered to be of low quality and does is not always directly link

to a spare part number and does not cover the whole population, the data is therefore considered to exist but incomplete. Giving it a sensitivity score seen table 16.

TABLE 16: THE CURRENT SUPPLY CHAIN VISIBILITY SCORECARD FOR LIFE LENGTH AT VG.

Parameter	Sensitivity	Accessibility	Intelligence	Decision Relevance	Fitness
Technical life length data	1/4	0/6	0/4	0/10	4 %

Life length sensor data
Finally, a relevant data parameter for determining the probability of failure is the "reliability given time" based on sensor data. This data has potential to create a more accurate understanding of demand since it is possible to continuously determine probability of failure. In practice meaning that the deviation of distribution will be very low for the probability of failure in short term. To have an updated "reliability given time" is currently in project phase and therefore very limited in the amount of spare parts. Sensor based reliability suffers from the fact that it is not always directly related to a spare part number but instead complete systems. This provides the basis of giving it a sensitive score as presented in table 17.

TABLE 17: THE CURRENT SUPPLY CHAIN VISIBILITY SCORECARD FOR LIFE LENGTH SENSOR DATA AT VG.

Parameter	Sensitivity	Accessibility	Intelligence	Decision Relevance	Fitness
Life length sensor data	1/4	0/6	0/4	0/10	4 %

5.2.3 BIAS DATA

Variable *K*, in equation 4, is used to adjust the demand for a bias that might occur when making estimations. The variable could be defined based on many different factors. For example, this could be due to customer loyalty, meaning that when a part breaks it is not sure that they will buy the spare part. Additionally, it could also be that the life data does not cover the entire population. Furthermore, demand or supply could be higher because of scrap in the aftermarket supply chain. Another example is that not all procurement is directly linked to a failure e.g. if the life length is generally set longer or shorter than the actual outcome. *K* could be adjusted continuously by cross checking with the real demand and adjust it over time when real demand data has been received. Currently at VG there is close to no data collection in any of the explained areas.

Customer loyalty data
When it comes to customer loyalty there are attempts to understand it, this focus more on the satisfaction, but there are no data regarding how many customers that choose alternative spare parts supplied by competitors. This means that no data are currently captured to support the variable, giving a sensitivity score of seen in table 18.

TABLE 18: THE CURRENT SUPPLY CHAIN VISIBILITY SCORECARD FOR CUSTOMER LOYALTY RELIABILITY AT VG.

Parameter	Sensitivity	Accessibility	Intelligence	Decision Relevance	Fitness
Customer loyalty data	0/4	0/6	0/4	0/10	0 %

5.2.4 STOCHASTIC VARIABLE

The stochastic variable ε, from equation 4, is defined as the variance in demand that occurs based on the nature of uncertainty for demand. This variable also includes all other factors that are not considered, since a prediction never completely reflects reality. Its value is determined by randomness created by differences of the actual demand and the predicted demand, $D(t)$. Since it's a stochastic variable it cannot be captured and therefore is not scored by the supply chain visibility scorecard matrix. The goal is to achieve a low variance of the variable. The lower the variance the higher the forecast accuracy, for a reliability based forecast.

5.3 DIFFERENT LEVELS IN SUPPLY CHAIN VISIBILITY

Through this study it has been identified that 4 different levels of supply chain visibility exists which support and disrupts current planning processes for the aftermarket. In order for VG to integrate big data in their current forecasting processes there has to be significant improvements in the transfer of data to the MM division, which was presented in the previous chapter. The integration of data strongly depends on the needed level of supply chain visibility. Where each level describes to what extent and in which situations big data can facilitate decision making. The different levels are presented in table 19.

TABLE 19: IDENTIFIED LEVELS OF SUPPLY CHAIN VISIBILITY FOR INTEGRATING BIG DATA IN FORECASTING.

Solution	Sensitivity	Accessibility	Intelligence	Decision Relevance	Fitness	Solution cost
Basic level	3/4	1/6	0/4	0/10	17 %	Low
Medium level	3/4	3/6	0/4	0/10	29 %	Low to medium
High level	4/4	4/6	2/4	2/10	50 %	Medium to high
Advanced level	4/4	5/6	4/4	10/10	95 %	High

Basic level
As shown in table 19 basic level defines a supply chain visibility where data correlating to improving forecasts is at the minimum level of integration, for facilitating the planning processes for spare parts. The data should be integrated and accessible for decision making even though the preferred method or system may not be used and there is no intelligence derived from the data, since the data can only be used for giving an indication for potential changes in the market. This level would rely on a high organizational willingness over a long period of time, since other stakeholders responsible for the data source would have to continuously share the information when requested. This level is good representation of where VG can strive for in the short term for implementing big data in forecasting. Because the data required is captured and integrated,

however for other purposes than forecasting. But with low costs data could be integrated in the planning process and improving forecasts could be done on ad-hoc basis.

Situations or operational activities in the planning process where such a level is relevant is when a spare part number becomes obsolete or the supply is decreasing. Because in such situations there often is a need to perform final purchase in order to fill inventory for future demand. However, considering that forecasting with a long time horizon is difficult, this basic level of supply chain visibility can facilitate and create an understanding of the current and future market situation, through big data. Furthermore, it provides a very basic tool for market understanding when replacing one obsolete spare part with several new ones, through access to such data a basic understanding of how expected demand changes is created. These examples suggest that a basic level of supply chain visibility limits the usage of big data because of the difficulties to extract and integrate the data. Such a level facilitates the planning process through only being useful for important or significant events, occurring a couple of times per year as described in the examples of operational activities above.

Medium level
According to table 19 supply chain visibility at a medium level is similar to the basic level with a significant difference in the extent of integration of captured data. In relation to the basic level a medium level moves from using ad-hoc methods to a standardized approach for integrating data. The captured data remains the same however it is how this data integrates with planning processes which improves. A higher level of integration provides an underlying tool for supporting forecasting, by providing a more feasible way to weekly or even daily extract data that correlates to the increase/decrease in demand of spare parts.

At this level of supply chain visibility several operational activities of the planning process for spare parts is affected, in addition to the activities improved at a basic level. A standardized method for integrating data provides a better basis for conducting analysis in order to find root cause of issues or deviations in demand. Because data which describe changes in the market, through population for example, gives a possibility of facilitating such analysis. Additionally, there are possibilities to improve the prediction of demand for spare parts which are newly introduced to the market. By having access to information which gives an indication of future demand one can more accurately set initial stock levels, in addition to having historic sales data for spare parts that are of similar nature.

High level
A high level of supply chain visibility, as viewed in table 19, captures all relevant data and is integrated to the extent that the decision maker can access it through any means, however the decision maker cannot easily change the approach for data integration. The most significant difference from basic and medium level is that data which is captured is utilized in order to create intelligence. Though somewhat basic, in the sense of only identifying when a decision is needed. In the planning process this means that less time is spent on analyzing deviations which is a necessity because of the large amount of SKUs which is forecasted. Furthermore, this level of supply chain visibility provides the first steps in automating data retrieval into demand planning for the aftermarket. It is used to support all regular activities, but consequently this depends on what the captured data is. Evidently, a high level provides the basic tools for supporting operational activities where the final decision is taken by the user, meaning that their expertise is necessary in order to make valid business decisions. A particular operational activity which is significantly impacted by this this level is the identification, analysis and changes made to seasonal demand behavior. Implementing data, which correlates to a specific demand pattern, in the planning process of spare parts has the potential for positively affecting the forecast and thus making it more accurate. Furthermore, spare parts classified as fast moving with high costs

represents the larger saving potential per part number in regard to decreasing both availability and tied up capital. An improvement in forecast accuracy would have a larger impact on such spare parts then slow and medium movers, which represents more SKU and requires a somewhat automated integration of data in the planning process in order for manual work to be feasible.

Advanced level
From table 19 the optimal supply chain visibility is described as advanced level where all aspects of the planning process for every single SKU is supported and improved through external data. Since there are several thousands of SKU, very low manual interaction is necessary. The significant difference from a high level of supply chain visibility is that the created intelligence is self-updating, meaning that the intelligence itself analysis when there is need for changing the approach for a making a business decision. Furthermore, the decision making is also conducted autonomously without involvement from humans except for debugging. In regards to operational activities this approach significantly facilitates forecasting for spare parts with a lumpy demand pattern. This is explained by the fact that lumpy demand is most difficult to predict and represent the majority of SKU in an aftermarket supply chain. An autonomously working supply chain visibility approach, which uses external data, can provide wider foundation of information in analyzing said demand. Which subsequently would have a bigger impact on spare parts classified as slow and medium movers. However, such an approach creates issues in trust where decision makers transfer all responsibility to a process. The risks associated with this can be mitigated through extensive iterative testing prior to implementing autonomy. Thus securing and safeguarding against predictable issues that might arise.

5.4 TRANSLATING BIG DATA INTO DEMAND

This chapter describes and elaborates around the sophistication level of combining different data sets in order to create demand pattern for spare parts, this is based on the “captured data” (equation 4 and chapter 5.2). Four levels of sophistication has been identified, each of which relates to differently advanced approaches for translating big data into a demand, *D(t).* All of which can be viewed as reliability based forecasting method. With such an approach it is critical to consider forecast accuracy. When using double exponential smoothing, commonly used with historical sales data, forecast accuracy becomes less accurate with longer planning horizons and more accurate for closer planning horizons. This mean that generally the less sophisticated levels, with lower forecast accuracy, can still compete with double exponential smoothing when having a long planning horizon. But when using reliability based forecasting for short planning horizons it tends to require a more sophisticated level. Table 20 show the different levels and the necessary data parameters in each level in order to achieve adequate approximation of demand .

TABLE 20: IDENTIFIED LEVELS FOR TRANSLATING BIG DATA INTO A DEMAND WITH NECESASSRY DATA PARAMETERS.

	Population	Time of usage	Mileage and runtime	Reliability given time	Service schedule	Sensor data	Solution cost
Basic level	X						Low
Medium level	X	X		X			Low to medium
High level	X		X	X			Medium to high
Advanced level	X				X	X	High

Basic level
The basic level, found in table 20, is based on having data regarding the population of a spare part number, variable *n*, and setting the probability of failure, probability *P,* equally for any period of time. Meaning that it doesn't use any life data for specifying percentage of failure. Instead there is a set percentage of failure, which is the same for the entire population. The probability of failure is based on a rectangular distribution. With this approach only population data is needed in order to give an estimate of demand, however with consistently updated data in order for the approach to remain fairly accurate. A basic level makes no distinction between the usage of spare parts, where two main categories can be service parts and breakdown parts. It is helpful for determining the demand for longer planning horizons. Even though it might spawn lower forecast accuracy it can be used for indicating trends of changes in demand and make rough estimates of demand, years ahead in time. Furthermore, it can be used to identify ramp ups by identifying when a population is increasing rapidly which consequently mean that the demand will eventually increase.

Medium level
From table 20 it can be viewed that this level provides an estimation of demand through combining data regarding the time of usage, when the product of which the spare part is part of was sold and the entire population. This gives rough estimate of when a demand can potentially occur. Such an approach needs no updated data in real time, such as sensor data, mileage or run-time. By simply knowing when a spare part was put into usage, which is interpreted as life data *x(t)*, the probability of failure, variable *P*, can be set to a probability distribution such as normal distribution or Weibull distribution. Where the mean value, μ, can be translated into expected life length in time (for breakdown parts) or service intervals in time (for service parts). The standard deviation, σ, can be estimated by the team working with the planning process for spare parts through analyzing similar or other spare parts. The bias variable, *K*, can be used as an adjustment factor for any bias in the life length data. This approach achieves the same results as a basic level approach but with a higher forecast accuracy, because it uses more datasets giving a more accurate probability of failure. It can, as a basic level approach, also give a better understanding of longer planning horizons, which are particularly appropriate for ramp ups but also during the declining phase of the life-cycle.

High level

A high level of sophistication for translating big data into demand, see table 20, would use more advanced datasets for the variable life data, *x(t)*. Data parameters presented in chapter 5.2 which are useful for this approach are mileage and runtime. By utilizing real-time data the market condition can be monitored more closely and changes in demand would be updated more frequently. The probability of failure, *P*, is determined in the same way as for a medium level approach. Furthermore, run time and mileage data over a period of time gives rise to the possibility of showing how the usage of spare parts are growing over time, which in effect gives more precise estimates for shorter planning horizons.

Advanced level

An approach at advanced level is defined, as seen in table 20, when demand can be estimated with a high accuracy and quality. This would be done through using sensor data as a basis of analyzing for when failure might occur. This gives rise to the preventive maintenance approach with a more proactive method for handling failure. In order for this to be feasible there is a need of combining said sensor data and its analysis with maintenance schedules in order to make sure that the material flow is following the demand. A distinction between service parts and breakdown parts is that the former would set the probability of failure, *P*, to one or zero. However, it is important to point out that the probability of failure, *P*, for breakdown parts should not be done within the department handing the planning process. Instead it should be done at departments which have more technological knowledge, with a self-interest in achieving a preventive maintenance approach in order to secure uptime of vehicles. In practice, this means that the captured data can be viewed as the probability of failure with intrinsic life data. Thus meaning that the department handling the planning process doesn't need to have supply chain visibility towards the life data, *x(t)*, which could be based on data parameters already presented (mileage, runtime or sensor data).

5.5 TWO DIMENSIONAL FRAMEWORK

The results of this thesis is a framework with two dimensions, presented in chapters 5.3 and 5.4, which can be considered as significant aspects when incorporating big data into forecasting for the aftermarket. Each of the two dimensions, sophistication and integration, go from basic to advanced level and the former is defined as the level of sophistication for translating big data into a demand. The latter dimension is defined as the level of supply chain visibility for integrating big data in the planning process of spare parts. Both of which has a considerable impact regarding feasibility of adopting big data in forecasting. The level of integration is dependent on the frequency and extent of spare part numbers in the planning process. Furthermore, the level of sophistication is dependent on the quality and accessibility of available data in the focal company, which in turn determines the forecast accuracy. From these two dimensions we conclude three stages of advancement (beginner, intermediate and expert) when incorporating big data into forecasting. An expressive figure of the developed framework can be viewed in figure 14.

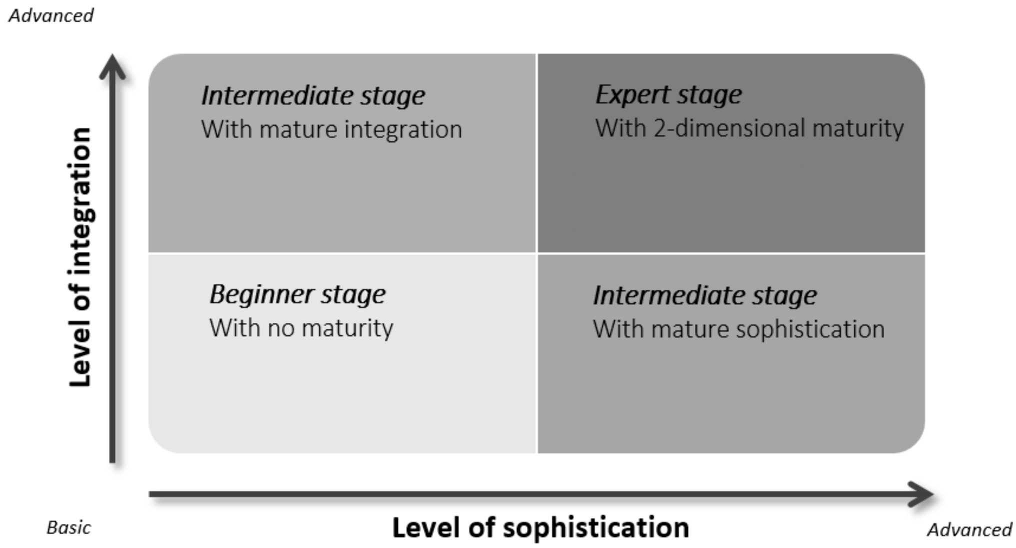

FIGURE 14: THE MATRIX DESCRIBES DIFFERENT STAGES IN ADOPTING BIG DATA IN FORECASTING. BASED ON TWO DIMENSIONS, LEVEL OF INTEGRATION AND SOPHISTICATION.

Beginner stage

As viewed in figure 14, this is the initial stage for automotive companies who strive for incorporating big data into forecasts. The fundamentals of this stage is to reach basic level in both dimensions. With a particular focus of identifying the population on the market and integrating this, on a basic level, with planning processes for spare parts.

Intermediate stage

From figure 14 it can be viewed that automotive companies can move towards expert stage in two different alignments. Either with a focus on sophistication or integration, which is effected by the maturity of the planning process at a company. However it is important to point out that companies have the possibly of moving from a beginner stage straight to an expert stage if enough resources are invested.

Expert stage

As seen in figure 14, this is the final and end stage for automotive companies looking for improving forecasts through big data. It is based on moving towards high automation in integrating big data into planning processes and also using highly sophisticated methods for translating big data into a demand.

6. CONCLUSION

Through this master thesis it has been concluded that big data has a great potential for improving forecast accuracy for the aftermarket. It was identified that data created downstream the aftermarket supply chain gives insight about real demand from end-customers. By combining data sources there is a possibility of short-circuiting information flow to support the planning process of spare parts and increase supply chain visibility. Furthermore, solely using historic sales data when forecasting in aftermarket supply chain which has multiple layers of warehousing creates significant bullwhip effects. Thus it was found that forecasts should not only be based on historic sales data but also use other sources of data which indicate market changes, as it has the potential for greatly improving forecast accuracy. Which consequently has positive effect on capital gains and customer satisfaction, through lower inventory and higher availability. Additionally, for sizable automotive companies there is a great importance of creating organizational willingness and cross-functional teams, in order to support the development of aftermarket forecasts through big data. The needed data is often owned by other stakeholders than the ones responsible for the aftermarkets material flow.

Furthermore, this thesis developed a framework which consider two dimensions of using big data with forecasts. Both of which are considered to have equally significant impact in the success of adopting big data. One dimension being the level of sophistication for translating big data into a demand and the other dimension being the level of supply chain visibility needed in order to support planning processes of spare parts.

Level of integration, from the framework presented in chapter 5.5, suggest that there are four different levels of supply chain visibility which support and disrupt planning process of spare parts when adopting big data in forecasting. These different levels are characterized by needed improvements in the transfer of data to the department responsible for material flow of the aftermarket. The basic level in this dimension is limited to mainly being useful for important or significant events which only occur a couple of times per year. This is due to difficulties for extracting and integrating big data in the current planning process. The medium level provides a more standardized approach and facilitates a more frequent decision making. In practice this means that it provides a better basis for conducting analysis which aims to find root cause of issues or deviations in demand. A high level of supply chain visibility support planning processes by automatically identifying when decisions are needed. At an advanced level the increase in automation better facilities the high amounts of SKU in an aftermarket supply chain. Thus spare parts classified as fast moving which has high cost represents the larger saving potential in terms of increasing availability and decreasing tied up capital.

Level of sophistication, from the framework presented in chapter 5.5, also suggest that there are four levels for of translating big data into a demand. This translation is based on creating a reliability based forecasting method, using equation 4 from the theoretical framework. Through utilizing life data this means that less sophisticated levels, with lower forecast accuracy, can still compete with time-series based forecasts when having a long planning horizon. When using reliability based forecasting for short planning horizons it requires a more sophisticated level, i.e. more advanced life data. Moving towards a more sophisticated level implies that automotive companies are moving towards proactive maintenance in contrast to reactive. Which greatly improves the uptime of vehicles and in effect also positively influences the material flow. However, in order to achieve the highest level of sophistication there is a need of greatly investing in developing knowledge and technological capabilities for measuring the status of the spare parts when in use. This type of life data was identified as sensor data and combining this with additional datasets provides a thorough understanding of demand for spare parts.

The developed two dimensional framework described three different maturity stages (beginner, intermediate and expert) for companies wanting to adopt big data into aftermarket forecasts, as viewed in figure 14. It was found that automated intelligence in regard to translating big data and integrating into planning processes was key for companies wanting to move towards a greater maturity in big data adoption. The level of integration strongly depended on the frequency of decision disruption and the extent of spare parts number in the planning process. The level of sophistication was identified to depend on the available data in the focal company. For companies

7. DISCUSSION

This section elaborates on strengths and weaknesses in the study. It proposes areas of interest for future research and describe contributions, both theoretical and practical.

Practical Contribution

This study has provided a framework for the aftermarket in the automotive industry. Which describes how to translate big data into a demand, as well as how to approach the visibility needed to disrupt decision in the planning process of spare parts. For VG this study has mapped all data parameters relevant for moving towards a reliability based forecasting method. Additionally, the quality of all the identified data parameters was evaluated. Further implication for VG is a roadmap regarding what to strive for in the area of big data in forecasting. Through the developed framework, automotive companies can identify their current state and thus start working towards the final stage of maturity in adopting big data into forecasting.

Theoretical Contribution

The theoretical contribution lies in the presented two dimensional framework. The framework and the different dimension are at its core explained in general terms. Simply put, if a company has big data correlating to failure of spare parts and share this data to the department responsible for aftermarket forecasting there is a possibility to improve the forecast accuracy. Implying that this framework has potential benefits for other industries and not only the automotive sector. For example, a lot of companies know from just sales data when a product was sold and therefore know how long the costumer has used the product. Knowing which spare parts are part of the product gives a possibility to estimate the demand for those spare parts.

Further research

The purpose of this study was to investigate the potential for using big data, i.e. life data, collected downstream the aftermarket supply chain in order to improve forecasts. The thesis was conducted during a period of 16 weeks and due to the time limit there was no possibility to study the monetary effects of big data implementation. Understanding which maturity stages of the two dimensional framework companies should strive for is based on a gain versus cost approach. There is little understanding of how big data usage at different maturity stages impact supply chain operations and business outcome, making this an important area for further research. Which is closely related to level of integration dimension in the framework. Benefits of receiving real time data from end-customers need to be weighed against the needed investments in infrastructure to support an advanced level of integration. Such decisions are highly depend on the focal company's aftermarket supply chain and thus must be further researched. At the same time benefits of real time data for supporting forecasts need significant investments in order to capture data, which effects the level of sophistication dimension in the framework. Again, resulting in a gain versus cost approach which needs to be further researched. Furthermore, the thesis time frame limited the possibility of evaluating improvements of forecast accuracy with quantitative methods. Which is mainly related to the level of sophistication. The study had little to no focus on evaluating equation 4 in a practical setting or implementation in current forecasting methods, resulting in a natural step for further research. It can be stated that an important area for further research regarding equation 4 is, how using different life data parameters for all spare parts classification and planning horizons effects the forecast accuracy. Furthermore, there was little exploration of how levels of sophistication, i.e. formula 4, works in combination with current or different forecasting methods. For example, the equation doesn't support seasonal effects meaning that any historical demand patterns are not considered if not the captured life data take it into consideration. Another aspect is that end-customer demand triggers demand upstream in an aftermarket supply chain, meaning that end-customer demand is delayed further upstream the supply chain. Subsequently this is important to consider and thus must be researched.

9. REFERENCES

Aberdeen Group, (2006). *The asset management benchmark report: moving toward zero downtime* [online]. Available at: http://diagsol.co.uk/wp-content/uploads/2007/12/aberdeenzerodowntime1.pdf [Accessed 22 Feb. 2016].

Aberdeen Group, (2012). *Supply Chain Visibility Excellence: Mastering Complexity and Landed Cost* [online]. Available at: http://www.feg.unesp.br/dpd/cegp/2012/LOG/Material%20-Complementar/Textos%20gerais/supply-chain-visibility%201.pdf [accessed 23 Mars 2016].

Accenture.com, (2014). *Global operations megatrends: bigdata* [online] Available at: https://www.accenture.com/t20160106T194441w/fi-en/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Digital_1/Accenture-Global-Operations-Megatrends-Study-Big-Data-Analytics-v2.pdf [Accessed 28 Jan. 2016].

Arribas-Bel, D. (2014) *Accidental, open and everywhere: emerging data sources for the understanding of cities.* Appl Geogr 49:45–53

Bacchetti, A. and Saccani, N. (2012). *Spare parts classification and demand forecasting for stock control: Investigating the gap between research and practice.* Omega, 40(6), pp.722-737.

Barratt, M. & Barratt, R., (2011). *Exploring internal and external supply chain linkages: Evidence from the field.* Journal of Operations Management, 29(5), 514-528.

Barton, D., and Court, D. (2012). *Making Advanced Analytics Work for You.* Harvard Business Review 90:79–83.

Bartwal, D., Skariah, A., Juyal, S., and Pundir, A. K. (2010). *Improving Aftermarket Supply Chain: A Top-Down Approach.* 2010 BPC World Conference and Exhibition. Mumbai.

Baudin, M. (2004). *Lean logistics*. New York, NY: Productivity Press.

Bryman, A. and Bell, E. (2015). *Business research methods*. Oxford: Oxford University Press.

Bundschuh, R. G. and Dezvane, T. M. (2003). *How to make after services pay off.* The McKinsey Quarterly, 4, 116-127.

Boylan, J.E. and Syntetos, A.A. (2008). *Forecasting for Inventory Management of Service Parts - Complex system maintenance handbook*. London: Springer.

Cassivi, L., Léger, P. and Hadaya, P. (2005). *Electronic commerce and supply chain integration: the case of the telecommunication equipment industry.* Business Process Mgmt Journal, 11(5), pp.559-572.

Cavalieri, S., Gaiardelli, P. and Ierace, S. (2007). A*ligning strategic profiles with operational metrics in after-sales service*. Int J Productivity & Perf Mgmt, 56(5/6), pp.436-455.

Cavalieri, S., Garetti, M., Macchi, M. and Pinto, R. (2008). *A decision-making framework for managing maintenance spare parts.* Production Planning & Control, 19(4), pp.379-396.

Chopra, S. and Meindl, P. (2013). *Supply chain management*. Boston: Pearson.

Cohen, M.A., Agrawal, N. and Agrawal, V. (2006). *Winning in the Aftermarket*. [online] Available at: https://hbr.org/2006/05/winning-in-the-aftermarket [Accessed 18 Mar. 2016].

Cohen, M. and Lee, H., (1990) *Out of Touch With Customer Needs? Spare Parts and After Sales Service*. Sloan Management Review; Winter 1990; 31, 2; ProQuest Central: pg. 55

Corbin, J. and Morse, J.M. (2003). *The Unstructured Interactive Interview: Issues of Reciprocity and Risks when Dealing with Sensitive Topics*. Qualitative Inquiry 9: 335–354.

Croston, J.D. (1972) *Forecasting and inventory control for intermittent demands*. Oper Res Q 23:289–303

Croxton, K. L., Garcia-Dastugue, S. J., Lambert, D. M., & Rogers, D. S. (2001). *The supply chain management processes.* The International Journal of Logistics Management, 12(2), 13-36.

Dubois, A. and Gadde, L. (2002). *Systematic combining: an abductive approach to case research.* Journal of Business Research, 55(7), pp.553-560.

Frowein, B., Lang, N., Schmieg, F. and Sticher, G. (2014). *Returning to Growth: A Look at the European Automotive Aftermarket* [Online]. Available at: https://www.acea.be/uploads/publications/BCG_Returning_to_Growth_Jul_2014.pdf [Accessed 23. Feb. 2016].

Gaiardelli, P., Saccani, N. and Songini, L. (2007). *Performance measurement systems in after-sales service: an integrated framework.* International Journal of Business Performance Management, 9(2), p.145.

Goffin, K. (1999). *Customer support: A cross-industry study of distribution channels and strategies.* International Journal of Physical Distribution & Logistics Management, 29(6), pp.374-398.

Goffin, K., and New, C. (2001). *Customer support and new product development-An exploratory study.* International Journal of Operations & Production Management, 21(3), 275-301.

Golicic, S., Davis, D., McCarthy, T. and Mentzer, J. (2002). *The impact of e-commerce on supply chain relationships.* International Journal of Physical Distribution & Logistics Management, 32(10), pp.851-871.

Handfield, R. and Nichols, E. (2002). *Supply chain redesign*. Upper Saddle River, NJ: Financial Times Prentice Hall.

Hassani, H. and Silva, E. (2015). *Forecasting with Big Data: A Review.* Ann. Data. Sci., 2(1), pp.5-19.

He, F. and Tao, T. (2014). *An improved coupling model of grey-system and multivariate linear regression for water consumption forecasting.* Polish Journal of Environmental Studies, *23*(4).

Heaney, B. (2013). *Supply chain visibility – A Critical Strategy to Optimize Cost and Service.* Aberdeen Group, 20.

Huiskonen, J. (2001). *Maintenance spare parts logistics: Special characteristics and strategic choices.* International journal of production economics, 71(1), 125-133.

Hyndman, R. and Koehler, A. (2006*). Another look at measures of forecast accuracy.* International Journal of Forecasting, 22(4), pp.679-688.

Jouni, P., Huiskonen, J. and Pirttilä, T. (2011). *Improving global spare parts distribution chain performance through part categorization: A case study.* International Journal of Production Economics, 133(1), pp.164-171.

Kennedy, W. J., Patterson, J. W., and Fredendall, L. D. (2002). *An overview of recent literature on spare parts inventories.* International Journal of production economics, 76(2), 201-215.

Lawrenson, J. (1986). *Effective Spares Management.* International Journal of Physical Distribution & Materials Management, 16(5), pp.3-111.

Lee, H., Padmanabhan, V. and Whang, S. (2015). *The bullwhip effect in supply chains.* IEEE Engineering Management Review, 43(2), pp.108-117.

Lohr, S. (2012). *The Age of Big Data.* NY Times. [online] Available at: http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html [Accessed 9 Mar. 2016].

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). *Big data: The next frontier for innovation, competition, and productivity.* McKinsey Global Institute

Marshall, C., & Rossman, G. B. (2014). *Designing qualitative research*. Sage publications.

Mayer-Schönberger, V. and Cukier, K. (2013). *Big data*. Boston: Houghton Mifflin Harcourt.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). *Big data. The management revolution.* Harvard Bus Rev, *90*(10), 61-67.

McIntire, J., 2014. *Supply chain visibility: from theory to practice.* Surrey: Gower.

Nelson, W. B . (2005). *Applied life data analysis* (Vol. 577). John Wiley & Sons.

Petrović, D. and Petrović, R. (1992). *SPARTA II: Further development in an expert system for advising on stocks of spare parts.* International Journal of Production Economics, 24(3), pp.291-300.

Phelan, A., Griffiths, J. and Fisher, S. (2000). *Pushing worldwide aftermarket support of manufactured goods.* Managing Service Quality, 10(3), pp.170-178.

Puntanen, S. (2010). *Linear Regression Analysis: Theory and Computing by Xin Yan, Xiao Gang Su*. International Statistical Review, 78(1), pp.144-144.

Riege, A. M. (2003). *Validity and reliability tests in case study research: a literature review with "hands-on" applications for each research phase.* Qualitative market research: An international journal, *6*(2), 75-86.

Robinson, A. (2014). *3 Automotive Aftermarket Trends in the Supply Chain*. [online] Transportation Management Company. Available at: http://cerasis.com/2014/10/20/automotive-aftermarket-trends/ [Accessed 25 Mar. 2016].

Romeijnders, W., Teunter, R. and van Jaarsveld, W. (2012). *A two-step method for forecasting spare parts demand using information on component repairs.* European Journal of Operational Research, 220(2), pp.386-393.

Saccani, N., Johansson, P. and Perona, M. (2007). *Configuring the after-sales service supply chain: A multiple case study.* International Journal of Production Economics, 110(1-2), pp.52-69.

Seuring, S. and Gold, S. (2012). *Conducting content-analysis based literature reviews in supply chain management.* Supply Chain Management: An International Journal, 17(5), pp.544-555.

Slater, P. (2010). *Smart inventory solutions.* New York, N.Y.: Industrial Press.

Shi Y (2014) *Big Data: history, current status, and challenges going forward.* Bridge 44(4):6–11

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). *Analytics: The real-world use of big data: How innovative enterprises extract value from uncertain data*. IBM Institute for Business Value.

Ton, M. (2013). *Big Data's Greatest Power: Predictive Analyticst*. [online] Available at: http://www.equest.com/cartoons/cartoons-2013/big-datas-greatest-power-predictive-analytics/ [Accessed 18 Feb. 2016].

Tucker, P. (2013). *My Predictions for 2014 and Beyond and Why They Might Not Come True.* [online] Slate Magazine. Available at: http://www.slate.com/articles/technology/future_tense/2013/10/futurist_magazine_s_predictions_on_quantum_computing_big_data_and_more.html [Accessed 24 Feb 2016].

Uschold, M. and Gruninger, M. (2004). *Ontologies and semantics for seamless connectivity.* ACM SIGMOD Record, 33(4), p.58.

Volvo Group, (2016A). *DIP GOT Presentation*. Unpublished internal document. Volvo Group.

Volvo Group, (2016B). *MM General presentation*. Unpublished internal document. Volvo Group.

Volvo Group, (2016C). *DIM Communication material*. Unpublished internal document. Volvo Group.

Volvo Group, (2016D). *Refill Presentatio*n. Unpublished internal document. Volvo Group.

Volvogroup.com, (2016a). *Parts Distribution - Parts Distribution : Volvo Group Global*. [online] Available at: http://www.volvogroup.com/group/global/en-gb/productsandservices/logisticssolutions/partsdistribution/Pages/default.aspx [Accessed 18 Mar. 2016].

Volvogroup.com, (2016b). *Guarddog Connect technology - Guarddog Connect technology : Volvo Group Global*. [online] Available at: http://www.volvogroup.com/group/global/en-gb/researchandtechnology/safetyanddriverenvironemt/GuarddogConnecttechnology/Pages/default.aspx [Accessed 18 Mar. 2016].

Volvogroup.com, (2016c). *About us - About us : Volvo Group Global*. [online] Available at: http://www.volvogroup.com/group/global/en-gb/volvo%20group/Pages/aboutus.aspx [Accessed 20 Apr. 2016].

Wagner, S. and Lindemann, E. (2008). *A case study-based analysis of spare parts management in the engineering industry.* Production Planning & Control, 19(4), pp.397-407.

Waller, M. and Fawcett, S. (2013). *Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management.* J Bus Logist, 34(2), pp.77-84.

Wang, W. and Syntetos, A. (2011). *Spare parts demand: Linking forecasting to equipment maintenance.* Transportation Research Part E: Logistics and Transportation Review, 47(6), pp.1194-1209.

Wirelesscar.com. (2016). *Our services*. [online] Available at: http://wirelesscar.com/our-services/ [Accessed 18 Mar. 2016].

World Economic Forum, (2012). *Big Data, Big Impact: New Possibilities for International Development.* [online] Available at: http://www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development [Accessed 28 Jan. 2016].

Yin, R. (2013). *Case study research*. Thousand Oaks, Calif.: Sage Publications.

Zikopoulos, P. (2012). *Understanding big data*. New York: McGraw-Hill.

APPENDIX A

A semi-structured interview form which was used when interviewing employees responsible for the planning processes of the aftermarkets material flow.

GENERAL QUESTIONS

- What department are you working in?
- What is your role in that department?
- Where in the aftermarket supply chain does your department control the material flow?
- What are your operational tasks (daily/weekly/monthly)?
- What are your long term strategic “tasks”?
- What improvements within forecasting are you working towards?
- What KPIs do you have and how are these measured?

INFORMATION FLOW

- What data does your department receive from other departments?
- What data is generated in your department?
- What data is sent out from your department?

MATERIAL FLOW

- What data is used as basis for your forecasts – both qualitative and quantitative?
- What data is used to initiate replenishment?
- What data initiates your demand?
- Who is your customer?
- What is your average lead time to customers?
- What is the total value of your tied up capital, on average?
- Which spare parts are difficult to forecast and why?
- What characterizes spare parts that reduce your overall forecast quality the most?

VIEW OF BIG DATA

- What data from other teams in your division could improve your forecast accuracy?
- What data from the entire organization could improve your forecast accuracy?
- What data outside the organization might improve your forecast accuracy?
- Which products have the greatest potential to improve your overall forecast accuracy?

APPENDIX B

A complete list of conducted interviews at VG.

Date	Department	Title	Topic
04.02.2016	GTO	MM concept development	Information and system mapping
04.02.2016	GTO	Demand and inventory planner	Planning process
04.02.2016	GTO	Demand and inventory planner	Planning process
05.02.2016	GTO	Demand and inventory planner	Planning process
08.02.2016	Volvo Trucks	Financial process specialist	Information and system mapping
09.02.2016	GTO	Dealer inventory manager	Planning process
09.02.2016	GTO	Refill analyst	Planning process
12.02.2016	Volvo Trucks	Manager connected services support	Information and system mapping
15.02.2016	GTO	Business analyst	Information and system mapping
16.02.2016	GTO	Director logistics service planning	Information and system mapping
17.02.2016	Volvo Buses	Manager maintenance and repair	Information and system mapping
22.02.2016	GTT	Systems engineer Function owner	Information and system mapping
24.02.2016	Volvo Trucks	Workshop product manager	Information and system mapping
01.03.2016	Volvo Trucks	Director of planning and portfolio management Project manager	Information and system mapping
01.03.2016	Volvo Trucks	Project manager connectivity	Information and system mapping
11.03.2016	GroupIT	BI Consultant Analyst IT	Information and system mapping
06.04.2016	Volvo Trucks	Service development specialist	Information and system mapping
27.04.2016	Volvo Trucks	Market analyst	Information and system mapping
27.04.2016	Volvo Trucks	Workshop product manager	Information and system mapping
10.05.2016	GTT	Project manager	Information and system mapping
10.05.2016	GTT	Business developer soft products	Information and system mapping

10.05.2016	GTT	Systems engineer Specialist uptime technology and on-board diagnostics	Information and system mapping
10.05.2016	GTT	Senior expert uptime and maintenance	Information and system mapping

APPENDIX C

Showing references used for defining terminology in the thesis work.

Domain	Subject	References
Supply chain management	Supply chain management	Chopra and Meindl (2013) Lee et al. (1997) Nicholas (2002) Wanger and Lindeman (2008)
	Logistics	Baudin (2004) Huiskonen (2001)
Aftermarket	Aftermarket	Bartwal et al. (2010) Bundschuh and Dezvane (2003) Cavalieri et al. (2007) Cohen and Lee (1990) Cohen et al. (2006) Gaiardelli et al. (2006) Goffin (1999) Goffin and New (2001) Phelan et al. (2000)
	Automotive	Frowein et al. (2014) Robinson (2014)
Spare parts	Spare parts	Bacchetti and Saccani (2012) Boylan and Syntetos (2008) Jouni et al. (2011) Kennedy et al. (2002) Lawrenson (1986) Petrović and Petrović (1992) Romeijnders and Jaarsveld (2012)
	Inventory	Slater (2010)
Big data	Big Data	Accenture (2014) Arribas-Bel (2014) Hassani and Silva (2015) Lohr (2012) Manyika et al. (2011) McAfee et al. (2012) Schroeck et al. (2012) Shi (2014) Tucker (2013) Uschold and Gruninger (2004) World Economic Forum (2012) Zikopoulos (2012)
	Connectivity	Uschold and Gruninger (2004)
	Predictive Analytics	Ton (2013) Waller and Fawcett (2013)
Forecasting	Forecasting	Croston (1972)
	Spare parts forecasting	Cassivi et al. (2005) Cavalieri et al. (2008) Petrovic and Petrovic (1992)
	Forecast accuracy	Hyndman and Koehler (2006)

Supply chain visibility	Supply chain visibility Relationships	Aberdeen Group (2006) Aberdeen Group (2012) Barratt and Barratt (2011) Croxton et al. (2001) Heaney (2013) McIntire (2014) Golicic et al. (2002)
Regression	Linear regression	He and Tao (2014) Puntanen (2012)
Life data analysis	Applied life data analysis	Nelson (2005)