Man-Hour Estimations in ETO

A case study involving the use of regression to estimate man-hours in an ETO environment

Aravindh Anand Alagamanna

Simarjit Singh Juneja

Master’s Programme in Industrial Management and Innovation

Masterprogram i industriell ledning och innovation
Abstract

Man-Hour Estimations in ETO

Aravindh Anand Alagamanna and Simarjit Singh Juneja

The competition in the manufacturing industry has never been higher. Owing to the technological changes and advancements in the market, readily available data is no longer a thing of the past. Numerous studies have discussed the impact of industry 4.0, digital transformation as well as better production planning methods in the manufacturing industry. The Mass-Manufacturing industry, in specific, has gained efficiency levels in production that were previously unimaginable. Industry 4.0 has been discussed as the ‘next big thing’ in the manufacturing context. In fact, it is seen as a necessity for manufacturing companies to stay competitive. However, efficient production planning methodologies are a preliminary requirement in order to successfully adopt the new manufacturing paradigms. The Engineering-to-order (ETO) industry is still widely unexplored by the academia ETO industries, barely have any production planning methodologies to rely on owing to their complex production processes and high reliance on manual-labour. Regression techniques have repeatedly been used in the production planning context. Considering its statistical prowess, it is no surprise that even the newer machine-learning techniques are based on regression. Considering its success in the mass-manufacturing industry for production planning, is it possible that its usage in the ETO industry might lead to the same results?

This thesis involves a case study that was performed at an electrical transformer manufacturing plant in Sweden. After understanding the several operations that are performed in the production process, regression techniques are employed to estimate man-hours. The results from the study reconfirm the statistical prowess of regression and show the possibility of using regression in order to estimate man-hours in the ETO industry. In addition, several factors that can affect successful adoption of this tool in the production planning context are discussed. It is hoped that this study will lay the foundation for better production planning methodologies for the ETO industries in the future which might subsequently result in more data-driven decision making rather than instincts.

Keywords: Regression, Production Planning, ETO, Data-Driven Decision Making
The technological changes that have occurred in the last two decades have not only impacted our daily life immensely, but these changes have also trickled down to numerous other areas that we might not witness every day. For somebody who’s not familiar with the manufacturing industry, the way this industry has changed over the two decades can be quite a surprising experience to witness. It is shocking, yet fascinating to see how numerous advances in technology have subsequently assisted in increased efficiency levels that have subsequently assisted the manufacturers to be more competitive for the market. In fact, ‘Industry 4.0’ has become the talk of the town since it deals with successful adoption of the said technologies in order to increase productivity levels even further and hence, is being deemed as a necessity for the manufacturers to sustain in today’s competitive world.

However, the manufacturing industry doesn’t simply consist of just one type of manufacturers. There are numerous types of manufacturing industries, Mass-Manufacturing and Engineering-to-order being two of them. Mass Manufacturing is exactly what it sounds like, an industry which manufactures products in ‘mass’ numbers. The Engineering-to-order industry, on the other hand, is the exact opposite. Every product is designed and created specifically for every customer depending on the needs. The mass-manufacturing has benefited greatly from the technological changes as these changes have subsequently assisted in better planning methodologies and hence, lower costs. The Engineering-to-order (ETO) industry has not benefited the same way due to the numerous complexities. Regression analysis has been one of major statistical techniques that has been used successfully in a number of different contexts, production planning in the mass manufacturing industry being one of them. Considering the statistical prowess and the flexibility regression techniques offer, it might be possible to use them in an ETO context for better planning and hence, increase the competitiveness of the industry.

To investigate how it can be used, a case study was performed at an electrical transformer manufacturing company in Sweden, which operates as an ETO company. The study was performed at the winding department of the company. Winding is an important part of any electrical transformer. The possibility of using regression-based techniques for estimating man-hours, which is a part of the production planning process, was explored.
ACKNOWLEDGEMENT

This master thesis was a case study conducted during the spring semester of 2020. We have received an immense support which led to the successful completion of this study on time amidst the challenging situation the world is facing right now. We appreciate all the employees at the case company for their enthusiastic support in this study and their immersive knowledge on the industry have made this study possible. We specially acknowledge Mikael Burlin, the production development manager and the external supervisor of this study for actively encouraging the employees to participate in this study especially in this tough period. And special thanks to Linda Strandberg and Jonny Andersson for their never-ending support throughout this study. We also thank Matias Urenda Moris, the internal supervisor of this study for the valuable comments during the seminars and ideas for further improvements of this study. Right from the beginning of this thesis, the comments and ideas were of great assistance in making this study better. Finally, we would like to thank our colleagues for their valuable feedback during the seminars which boosted this thesis’s quality.

We would also like to mention that this study was equally contributed by both the authors. We have mutually agreed on all the topics that have been written in this report.

Aravindh Anand Alagamanna and Simarjit Singh Juneja

Uppsala, 14th October 2020
Table of Contents

1 INTRODUCTION .............................................................................................................. 1
1.1 Background .................................................................................................................. 1
1.2 Problematization ........................................................................................................... 2
1.3 Purpose .......................................................................................................................... 3
1.4 Research Questions ....................................................................................................... 3
1.5 Delimitations .................................................................................................................. 3
1.6 Structure of the thesis ................................................................................................. 4

2 LITERATURE REVIEW .................................................................................................. 5
2.1 ETO Environments ....................................................................................................... 5
2.2 Comparison with Mass Production .............................................................................. 5
2.3 Production Planning in ETO ....................................................................................... 6
2.4 Bidding in ETO ............................................................................................................. 8
2.5 Man-hours .................................................................................................................... 9
2.6 Historic data in Production Planning ......................................................................... 11
2.7 Data-Driven Analytics ............................................................................................... 12
2.8 Regression Analysis ................................................................................................... 13
2.8.1 Multiple Linear Regression Analysis ...................................................................... 13
2.8.2 Data in Regression ................................................................................................ 14
2.8.3 Output Analysis in Regression .............................................................................. 16
2.9 Statistical Software for Regression Analysis ............................................................. 17
2.10 Employee Acceptance ............................................................................................. 17
2.11 Standardization Programs in ETO .......................................................................... 18

3 METHODOLOGY ............................................................................................................ 19
3.1 Research Methodology ............................................................................................... 19
3.1.1 Quantitative Methodologies .................................................................................. 19
3.1.2 Qualitative Methodologies ................................................................................... 20
3.2 Research Design ......................................................................................................... 20
3.3 Research Approach .................................................................................................... 21
3.4 Data Collection ........................................................................................................... 22
3.4.1 Literature Review .................................................................................................. 24
3.4.2 Quantitative Data Collection ............................................................................... 25
3.4.3 Qualitative Data Collection .................................................................................. 25
3.5 Data Analysis .............................................................................................................. 27
3.6 Generalizability of this Research ............................................................................. 28
3.7 Reliability and Replicability ....................................................................................... 29
3.8 Validity ....................................................................................................................... 29
3.9 Ethics......................................................................................................................29

4 **EMPIRICAL DATA** ...............................................................................................32
4.1 Plant Description................................................................................................32
4.2 Product Description............................................................................................33
4.3 Constructing Database for Regression Analysis....................................................34
  4.3.1 Data-Collection methods.............................................................................34
  4.3.2 Detailed Engineering Plans..........................................................................34
  4.3.3 Detailed Production Plans...........................................................................35
4.4 Data Quality .......................................................................................................36
4.5 Data Cleansing ....................................................................................................38
  4.5.1 Missing Values ...........................................................................................38
  4.5.2 Outlier Detection .........................................................................................39
  4.5.3 Duplicate Values ..........................................................................................40
  4.5.4 Dimensionality Reduction ..........................................................................40
4.6 Construction of Regression Model......................................................................40
  4.6.1 R-Square Values ........................................................................................41
  4.6.2 Values of Estimated Coefficients ................................................................41
  4.6.3 Variance Inflation Factor (VIF) ..................................................................41
4.7 Approach 1: Ordinary Least Square Method ......................................................42
4.8 Approach 2: Non-Negative Least Square (NNLS)................................................43
4.9 Approach 3: Principal Components Regression (PCR) ......................................44
4.10 Approach 4: Average Values and Reduced Variance ......................................45
4.11 Approach 5: Increased Sample Size and Dimensionality Reduction ..............45
4.12 Validation ..........................................................................................................47
4.13 Data-Driven Decision-Making in ETO ..............................................................49

5 **ANALYSIS** .........................................................................................................51
5.1 Research Question 1 ..........................................................................................51
5.2 Research Question 2 ..........................................................................................56

6 **DISCUSSION** ...................................................................................................60
6.1 Data-Collection ..................................................................................................60
6.2 Union-Management relationship and Leadership needed in ETO .....................60
6.3 Union-Management Trust ................................................................................61
6.4 Should ETO industries aim for complete data-driven decision-making? ..........62
6.5 Process Innovation .............................................................................................62
6.6 Challenges during the study ..............................................................................64

7 **CONCLUSION** .................................................................................................65
7.1 Concluding the study .........................................................................................65
7.2 Academic and Practical Contributions ...............................................................66
7.3 Further Research ...............................................................................................66
REFERENCES .............................................................................................................................................. 68
Appendix A: Results from MiniTab using Ordinary Least Square Method ........................................... 78
Appendix B: Results from R using NNLS method ..................................................................................... 79
Appendix C: Eigen analysis of the Correlation Matrix ............................................................................. 80
Appendix C: Eigen Vectors of nine principal components ....................................................................... 81
Appendix C: PCR Results from MiniTab ................................................................................................. 82
Appendix D: Results from MiniTab through average inputs ..................................................................... 83
Appendix E: Results from MiniTab by combining Ordinary Least Square method and
Qualitative feedback ................................................................................................................................. 84
Appendix F: Interview Guide .................................................................................................................... 85
List of Tables

Table 1: Data sources used to answer the research questions .............................................. 21

Table 2: Designation of interviewee and number of interviews conducted with every interviewee ........................................................................................................................................ 27

Table 3: Summary of the case company .................................................................................. 33

Table 4: Classification of projects and sub-projects ................................................................. 33

Table 5: Representation of Detailed Engineering Plans template for various windings......... 34

Table 6: Representation of Detailed Production Plans template for various windings......... 35

Table 7: Preliminary database template .................................................................................. 36

Table 8: Validation of the resulted regression model against recently completed projects in the case company ........................................................................................................ 48
List of Figures

Figure 1: Explanatory Sequential Design................................................................. 19

Figure 2: Various types of data sources employed for this study ................................. 24

Figure 3: Structured Data-cleaning approach, as explained by Corrales, Corrales and Ledezma (2018) ...................................................................................................................... 38
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDD</td>
<td>Data driven decision making</td>
</tr>
<tr>
<td>ETO</td>
<td>Engineering-to-Order</td>
</tr>
<tr>
<td>MH</td>
<td>Man-hours</td>
</tr>
<tr>
<td>MS</td>
<td>Microsoft</td>
</tr>
<tr>
<td>MTS</td>
<td>Make-to-Stock</td>
</tr>
<tr>
<td>ML</td>
<td>Machine Learning</td>
</tr>
<tr>
<td>NNLS</td>
<td>Non-Negative Least Square</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PCR</td>
<td>Principal Component Regression</td>
</tr>
<tr>
<td>PPC</td>
<td>Production Planning and Control</td>
</tr>
<tr>
<td>RFP</td>
<td>Request for Proposal</td>
</tr>
<tr>
<td>RQ</td>
<td>Research Question</td>
</tr>
<tr>
<td>SME</td>
<td>Subject Matter Experts</td>
</tr>
<tr>
<td>VIF</td>
<td>Variance Inflation Factor</td>
</tr>
</tbody>
</table>
1 INTRODUCTION

This section will introduce the background of this research followed by the problematization. On the later part, purpose of this research will be explained with the research questions of this study. This section ends with delimitations and structure of the thesis that would assist in navigating through the study.

1.1 Background

Ever since the world was introduced to the concept of ‘industrialization’, there have been numerous technological leaps that have subsequently led to a change in the paradigms not just in the context of manufacturing, but the society too. These ‘leaps’ are known as the first industrial revolution (the introduction of machines), the second industrial revolution (usage of electricity) and the third industrial revolution (digitalization) (Lasi et al., 2014). With this advanced digitalization, the fourth industrial revolution is envisioned as a future that will be built on the foundation of the third industrial revolution and will result in highly efficient automated manufacturing systems. The combination of internet with the augmentation of ‘smart’ technologies is the culmination of industry 4.0, or the fourth industrial revolution (Lasi et al., 2014; Dallasega, 2018). In fact, investing in industry 4.0 has been discussed as extremely important for manufacturers in today’s world to remain competitive (Bosman, Hartman and Sutherland, 2019). However, in addition to the manufacturing industry, industry 4.0 is also expected to impact the ETO industry greatly (Dallasega, 2018).

But what is the ETO industry? With the increasing competition in the market, many manufacturers are now aiming to please customers by providing customized products specifically made to satisfy their requirements. ETO or Engineering-to-Order development has helped numerous companies serve customers in cases where the demands vary frequently. This kind of development involves designs that are specifically tailored-fit for each customer (Levandowski, Jiao and Johannesson, 2015). For companies involved in the ETO paradigm of development, every order is ‘engineered’ specific to the customer requirement. This newly engineered design is then produced and delivered to the customer (Rauch, Dallasega and Matt, 2018). Therefore, it is evident that ETO development can be highly complex.

There are expectations that industry 4.0 is going to transform the ETO industry (Dallasega, 2018). In fact, the wave of industry 4.0 is expected to enable the possibility of more customized products while also maintaining profitability (Ustundag and Cevikcan, 2017). However, before
the successful adoption of industry 4.0, accurate planning and resource allocation has been
determined crucial (ibid). In fact, Lasi et al. (2014) discussed how for a successful industry 4.0
implementation, there is first a need for innovative methodologies for planning manufacturing
systems. Unfortunately, as this study will discuss subsequently in-depth, there is a severe lack
of production planning methodologies for the ETO industry.

1.2 Problematization

The ETO environment has received less attention in research as compared to the ‘Mass
Production’ environment (Carvalho, Oliveira and Scavarda, 2015). Despite some
methodologies being developed for production planning in the ETO industry (Hendry, Amaro
and Kingsman, 1999; Spring and Dalrymple, 2000; Little et al., 2000), most of them are too
general and lack specificities that are required for an efficient production planning in an ETO
environment (Adrodegari et al., 2015). Considering how better production planning
methodologies are being considered as a preliminary requirement on the road to industry 4.0
(Lasi et al., 2014; Bendul and Blunck, 2019), it can be concluded that poor methodologies may
act as an obstacle for the ETO industry to progress towards industry 4.0. In addition, the lack
of suitable and efficient methodologies can also affect the competitiveness of an ETO company
(Stevenson, Hendry and Kingsman, 2005). As of the current state, in the ship building industry
(an example of the ETO industry), most of the man-hour estimations which is a significant
factor while planning a production are made by experts on-site based on their instincts (Hur et
al., 2015), which can be the case for most ETO companies.

The mass production industry employs machine learning techniques that are based on
regression methods (Lingitz et al., 2018), for data processing and predictions, showcasing how
effective regression methods can be in terms of processing data to enable accurate predictions.
In such a case, adopting regression techniques might be useful in building a production planning
tool. This could help ETO companies base their decisions regarding production planning on
data rather than instincts. It might subsequently reduce the errors between prediction and real-
life scenario as well as serve as a primary step towards building a robust production planning
methodology.
1.3 Purpose

The purpose of this study is to investigate how regression can aid in implementing data-based prediction (for man-hours) as well as identify the key factors that can contribute to successful adoption of it in a manual-labour intensive ETO company.

1.4 Research Questions

This thesis will focus on how regression analysis can be used on the production data from the shop floor. The first research question will address on how the ETO companies can make use of a popular statistical tool, regression analysis, on their production data to accurately estimate the man-hours for their future projects. The second research question will explore the factors that contribute for the successful adoption of the proposed model in the ETO company.

*Research Question 1*

How can regression analysis be used to estimate man-hours in an Engineering-to-Order environment?

*Research Question 2*

How can data driven production planning tools such as regression analysis be successfully adopted in ETO industries?

1.5 Delimitations

Feasibility is an important aspect of a study in order to ensure quality (Bryman and Bell, 2011). Considering the availability of time and other resources for this study, it is important to define the scope at the very beginning in order to ensure quality results. There are numerous statistical tools available that can be looked at and analysed to see if they would be suitable for planning purposes in the ETO environment. However, it is not feasible to do so. Regression has repeatedly been proved as an extremely versatile statistical tool that has been used to analyse numerous scenarios. In addition, it has also been proven quite effective through its usage in the mass manufacturing industry as well as in machine-learning techniques. Therefore, owing to its strong foundation, regression was the statistical tool of choice for this study. The thesis will only focus on the application of regression analysis specifically linear regression analysis in an ETO company and how it can be used to estimate the man-hours. And, this research is a single case study and will not focus on other ETO industries such as construction and aircraft industries due to time constraint. The name of the firm, interview participants, the projects and
the production parameters will not be revealed due to the confidentiality agreement. The interviews were not transcribed, but the information attained from the interviews was analysed to answer the research questions.

1.6 Structure of the thesis

The theoretical framework on which the thesis relies upon is discussed in section two, the literature review. This section comprises of theories and concepts that are discussed after reviewing various literatures on the related field of study. Section three constitutes the methodology where research approach, research design and data-collection techniques have been discussed that were subsequently used to answer the research questions. Section four explains the data collection methods with quantitative results. The results were then analysed to answer the research questions in Section five. Sections six follows with a generalized discussion on the results attained from this study. Finally, Section seven concludes the thesis along with the academic and practical contributions and further scope of this study.
2 LITERATURE REVIEW

This section explored the current literature on the relevant topics for this study. Published sources were critically analysed on areas such as production planning, ETO environments, regression analysis etc. and a detailed review has been presented.

2.1 ETO Environments

In the last few years, there has been quite an increase in the demand for the ‘customized’ products in the industrial market. In fact, it was pointed out as far back as 1972 that the market is moving towards a more customized approach (Ashour and Vaswani, 1972). But this phenomenon has in turn lead to the market becoming more and more fragmented (ibid). With this new type of demand the manufacturers have had to adopt new models in order to keep up with the market. Engineering-to-Order (ETO) is one of the models which is now being used by numerous manufacturers in multiple industrial sectors. In fact, in 2005, it was pointed out that the firms following an ETO model comprise of almost a quarter of all manufacturing in the Northern American continent (Cutler, 2005).

So, what is an ETO model? In the case of ETO models, the company has to newly design their product for every single order (Rahman Abdul Rahim and Shariff Nabi Baksh, 2003). This is done to conform to the customer’s specific requirements. According to Hicks and Braiden (2000), Engineering-to-Order (ETO) suppliers contribute significantly to the world’s economy. ETO industries despite being under the same ‘umbrella’, called ETO model, can still vary significantly in terms of the complexity and specificity of the product as well as the process (Bertrand and Muntslag, 1993). The production processes are highly customized to meet individual customer requirements and are produced usually in low volumes (Hicks and Braiden, 2000) and this low volume production with customized product specifications drives complexity (Grabenstetter and Usher, 2015).

2.2 Comparison with Mass Production

ETO’s operating strategies differ significantly from other prevailing production approaches. Therefore, to understand ETO better, it might be beneficial to contrast it to one of the most common manufacturing strategies i.e. mass production by identifying the differences (Adrodegari et al., 2015). As discussed, products in the ETO environment are identified as highly customized and in most cases, also non-repetitive (Hendry, Amaro and Kingsman, 1999;
Pandit and Zhu, 2007). What this means simply is that for almost every order, the company has to ‘engineer’ i.e. design a product from the scratch as per the customer’s requirements. Usually, an ETO company will have to incorporate both, the physical stage as well as the non-physical stage for every order it receives (Bertrand and Muntslag, 1993). The non-physical stage involves the design engineering of the product (Hendry, Amaro and Kingsman, 1999; Wikner and Rudberg, 2005; Gosling, M. and Naim, 2009) whereas the physical stage includes the manufacturing and assembly (Adrodegari et al., 2015). While, a mass manufacturer, on the other hand, will only have to incorporate the non-physical stage during the first order and only physical stages for the orders following the first order. Therefore, it is quite evident that ETO industries must deal with a lot of uncertainties. This uncertainty finds its roots in not only the configuration of the product, but the market itself is volatile with drastic changes occurring frequently (Adrodegari et al., 2015). In fact, a study pointed out that in the ETO industry, the number of orders and the product shipments change by almost 50% in terms of volume from year to year (Anderson, Fine and Parker, 2000). But it is also important to point that this might not be the case in all the ETO industries and can depend on the type of product being manufactured. However, the volatility of the ETO market is unequivocal. Considering the high degree to which the products manufactured in the ETO industry are customized, the outputs in the ETO industry are also called OKP or ‘One-of-a-kind’ production. This is as opposed to the mass manufacturing industry where the products are undifferentiable (Adrodegari et al., 2015). Considering the differences, the production planning frameworks in the ETO industry must also adapt to the conditions. Unfortunately, the ETO industry severely lacks in a Production Planning and Control process (Stevenson, Hendry and Kingsman, 2005).

2.3 Production Planning in ETO

Production Planning and Control (PPC) tools are essential for any manufacturing industry to meet the customer requirements in a highly competitive market (Stevenson, Hendry and Kingsman, 2005). A definition by Business Dictionary as cited in Kiran (2019, p. 7) is “Production planning is the administrative process that takes place within a manufacturing business and which involves making sure that sufficient raw materials, staff and other necessary items are procured and ready to create finished products according to the schedule specified”. Hence, production planning is a method of visualizing the production processes prior to the actual operations and making the decisions regarding resources based on facts. The role of an accurate PPC tool will be to accurately forecast the resources, delivery dates and competitive pricing (Stevenson, Hendry and Kingsman, 2005). Production planning process, plans and
controls the material supply and processing activities of an enterprise in such a way that labour force and the capital available are exploited effectively (Kiran, 2019).

As discussed, an ETO company’s supply chain consists of non-physical stages which are project tendering, product engineering & design, process planning activities etc. and physical stages which are manufacturing, assembly & installation (Hicks and Braiden, 2000). As discussed, supply chain process in these environments are characterized by high levels of uncertainty in specifications, demand, lead times and the duration of the processes (ibid). Even the market is quite unstable, making the entire value chain process to be volatile. Thus, to tackle all these uncertainties in the future, the ETO industries must plan their resources and capital in order to avoid significant losses. Moreover, an interesting point to note, Hicks and Braiden (2000) interviewing experience while conducting their study on how simulation can improve planning in ETO industries revealed that, managers in ETO industries considered resource utilization and production output measured in terms of standard hours.

Make-to-stock industries or Mass manufacturers rely on structured production planning tools for mass manufacturing on a daily basis, but these tools and frameworks are unsuitable to ETO industries as they do not acknowledge the fact that different projects are carried out at different times and the requirements can be changed mid-way (Rahman Abdul Rahim and Shariff Nabi Baksh, 2003). Therefore, using tools that are more suited to the traditional ‘Make-to-stock’ industry may not give the best results when used in the ETO context if they are not modified to suit the environment (ibid). Only a few methodologies are provided by the literature for managing production projects in an ETO industry (Hendry, Amaro and Kingsman, 1999; Spring and Dalrymple, 2000). The Supply Chain Operations Reference Model (SCOR 2010) is considered one of the first point of references for an ETO company (Adrodegari et al., 2015). This framework focuses on decision-making in the supply chain context for an ETO company (ibid). Another framework proposed by Little et al. (2000), discussed the process of scheduling in an ETO industry by including six sub-processes. As effective as it is at including the major activities that take place in an ETO company, it still requires a substantial amount of customization in order to be used in the different ETO industries (Adrodegari et al., 2015). There are also few other frameworks that discusses production planning and control in ETO industries. However, a study by Adrodegari et al. (2015) reviewed these methodologies included in almost all the cases and concluded that the frameworks are either too general or they lack the specificity that an ETO company might require.
Thus, considering the volatility of ETO environment, complexities in products specifications, market dynamics and high competitiveness, planning the necessary resources with a structured approach becomes quintessential which would eventually lead to an accurate price quotation during project bidding stage.

2.4 Bidding in ETO

In addition to the factors that emphasize on production planning in ETO, one of the major factors encouraging the importance of planning in ETO environments, is bidding. In ETO environments, the customer usually chooses a supplier or a contractor on the basis of competitive bidding (Rothkopf and Harstad, 1994). A ‘Request for Proposal’ (RFP) is usually prepared by the customer which is then forwarded on to the potential suppliers that have the capability to supply this demand (Ishii and Muraki, 2011). Once the customer receives the required information, an analysis takes place amongst the existing proposals over several different parameters such as price, performance, reputation, time, method of delivery etc. (ibid).

But, how exactly does the bidding procedure work for a contractor? Wang, Xu and Li (2009) looked into the Chinese construction industry to understand how a contractor selects projects to bid on and how this procedure unfolds in the ETO environment. The study concluded that the selection in terms of which projects to bid on was an extremely complicated process. Numerous factors such as the market situation, the availability of the raw materials and even the probability of a successful outcome can affect whether a contractor chooses to bid or not. This is because the contractor can attract a penalty if they fail to provide the outcome that was promised earlier (Rothkopf and Harstad, 1994). And products manufactured by ETO companies are often used in large projects for instance construction, ship building etc, and hence it is usual that customers ought to impose very high penalty charges for lateness (Grabenstetter and Usher, 2015).

The clients select projects (or contractors) on the basis of cost, if all the other parameters are mostly similar in the proposals. Therefore, it increases the competition for the different ETO contractors even further in not only providing the right technological solution, but also estimating the cost of the project in a correct manner. If the cost estimates are too high, the proposed cost of the solution will be high and there will be a higher probability that the competitors will outbid the proposal. But, if the cost estimations are too low, despite winning the contract, the contractor might still incur a loss (Ishii and Muraki, 2011). Sometimes, the quickness in responding to RFP will also be evaluated which makes the customer believe in the
experience and the responsiveness of the supplier. Therefore, it can be said that accurate costing techniques is of utmost importance in ETO industries considering the different intricacies of competitive bidding. There are some methodologies that have been developed in order to help ETO industries correctly estimate their costs (Lunghi, Botarelli and Brizioli, 2007). One of the important parameters that is required to be estimated correctly for an accurate cost estimation is the close approximation of the skilled ‘Man-hours’ or ‘MH’ required for particular project (Ishii and Muraki, 2011). Unfortunately, despite being such an important factor to be considered in cost estimation, a lot of estimation models assume the number of skilled Man-hours to be unlimited, which is never the case in a practical situation (Ishii, Takano and Muraki, 2014).

Therefore, summarizing the information above, competitive bidding is an inherent part of ETO industries. In order to increase the chances of a successful bid, cost estimates play an extremely important role for the contractor. Accounting for ‘Man-hours’ is important in ensuring good cost estimates, however, it is a factor quite commonly ignored in many cost-estimation models. Therefore, estimating ‘Man-hours’ correctly can lead to more accurate results in estimating costs, which in turn can help in outcompeting other competitors in the bidding process.

2.5 Man-hours

All of the aforementioned information is somehow related or directly points to the importance of estimating man-hours of operations that a product undergoes to finally reach the customer. What are ‘Man-hours’? Taking an example of a simple machining operation, if a steel sheet needs to be cut into two equal parts and it takes two people a total of one hour to do the job, the total number of man-hours in this case is $2 \times 1 = 2$ Man hours (Hur et al., 2015). Having a look at similar ETO industries might help in understanding why there is such a strong correlation between accurate cost estimations and man-hour estimations. The shipbuilding industry is an example of a complicated ETO industry. Typically, in the ship building industry, the man hours have been estimated by experts on site (Hur et al., 2015). Unfortunately, these estimates by the experts are usually incorrect and lead to significant errors. The reason why man-hours are given such importance is because in the ship-building industry, the human labour costs account for as much as 50% of the production cost (ibid). However, it should be noted that this might not be the situation for all the companies under ETO umbrella. But it still underlies the importance of accurate man-hour estimations in the ETO industry which has relatively lower levels of automation (Mei et al., 2016) and hence, significant labour costs.
A study performed at a shipbuilding yard in Malaysia used Work breakdown structure in order to simplify the production process and in turn, estimate the man-hours more accurately. This approach was combined with the usage of historical data (Wan Abd Rahman, Mhd Zaki and Abu Husain, 2019). This innovative way of man-hours estimation resulted in simplifying the planning of a complicated project. In addition, the authors also concluded that using this approach might result in a better prediction regarding the cost as well as the schedule (Wan Abd Rahman, Mhd Zaki and Abu Husain, 2019). In fact, in addition to being used for cost estimations, knowing the man-hours data of different products also assists in production planning (Liu and Jiang, 2005) and this information can help in framing a basic production plan (Ye, Cui and Zhou, 2017).

This leads to the question, is there a way to accurately determine man-hours for a project? Some methodologies have been developed to estimate the man-hours required for a product (Liu and Jiang, 2005). The study by Liu and Jiang (2005) discusses the methodologies to estimate man-hours in an ETO industry that focuses on the ship-building industry. The ship-building industry usually follows a ‘fixed’ layout (Liu, Meng and Liu, 2013). Since the product itself in this industry is quite big and difficult to move, the ‘location’ of the product is fixed, and operations are carried out at this specific location (ibid). The methodologies discussed by Liu and Jiang (2005) calculate the man-hours based on the physical dimensions of the product such as area and weight as the material needs to be transported to the assembly site in order to carry out the production operations. While this methodology may work for industries following a ‘fixed’ layout, just calculating man-hours on the basis of the area as well as the weight of the final product may not work for all the industries. Therefore, as Adrodegari et al. (2015) pointed out, the frameworks that have been discussed in the literature for planning in ETO lack some aspects and they need to be customized for every specific industry in the ETO environment.

To summarize, this section discussed the relationship between three aspects of an ETO industry- Production Planning, Cost Estimations and Man-hour Estimations (or allocations). Now, it is quite evident that there is a strong correlation between the three and it is important for any ETO company to acknowledge it in order to remain competitive and profitable. It should be noted that using well-designed framework and methodologies might lead to an increase in the accuracy of man-hour estimations and subsequently, an increase in the accuracy of cost estimations in addition to better production planning (Liu and Jiang, 2005; Mei et al., 2016; Hur et al., 2015). Therefore, it can be argued that robust man-hour estimations can help
companies operating in the ETO environment take a step towards building better production planning methodologies as well as better cost estimations.

2.6 Historic data in Production Planning

Lingitz et al. (2018) have conducted a study at a semiconductor manufacturing site which is a make-to-stock industry where conventional planning often calculates average lead times which is based on their historical data. But this plan was unable to incorporate various details such as complexity in manufacturing, multiple routings and demands to high process resource efficiency. Hence, Lingitz et al., (2018) have recommended the use of Machine Learning (ML) techniques to be employed for accurate prediction of lead times. According to Alpaydin (2020), “machine learning is programming computers to optimize a performance criterion using example data or past experience”. According to Lingitz et al., (2018), the machine learning techniques are built on regression models and the input for the regression models is the historic data about their past projects which is obtained from manufacturing execution systems.

So, how is the historic data being collected and recorded? In make-to-stock sectors, the production involves a lot of automatic or semi-automatic processes, and data can be extracted if all the machines are properly integrated and hence data collection will be a simple task (Lingitz et al., 2018). In a study by B A, P N and P M (2020), cyber-physical system were incorporated into the shop floor for live monitoring of production data without much manual intervention. The study was conducted in a Small and Medium Enterprise company in India which manufactures plug shells for automobiles. Various sensors were used, in the machine which automatically sends the data to the cloud server from where the data can be extracted and used. While in another study by Zhong et al. (2013) in a large scale and heavy duty machineries industries, they had used RFID (Radio Frequency Identification) enabled real time manufacturing execution systems. Meanwhile, in ETO industries relying on manual labour use manual data acquisition techniques (Kumar and Shinde, 2019). Hence, the data collection process becomes very sensitive and the accuracy varies between people.

However, what is interesting is the fact when an ETO company is characterized by the non-standardized, customized products (Adrodegari et al., 2015), is it still possible to use historical data for planning for the future? Wouldn’t every operation that a newly designed product must go through be different? With different operating times? Interestingly, the literature points out that a lot of ETO companies are opting for a technique called ‘Mass-Customization’ (Gosling, M. and Naim, 2009; Willner et al., 2016; Wikner and Rudberg, 2005). In addition to this, ETO
companies also opt for strategies such as ‘Product Platforms’ and ‘Product Modularization’ (Johnsen and Hvam, 2019). Following the aforementioned methodologies enables ETO companies to reuse product designs and ‘modules’ to fulfil an order (Johnsen and Hvam, 2019). Despite different models, the modules and designs remain unchanged sometimes. They might just be ‘arranged’ differently. For example, two product modules might be used in product B and three of the same product modules might be used in product C. But the amount of time it takes to assemble one of the modules might not be significantly different. Therefore, studying the viability of using historical data for man-hour estimations in an ETO company while keeping the aforementioned information in mind might help in starting to build a foundation for better, data-driven planning and estimations in the ETO environment.

2.7 Data-Driven Analytics

Data-driven decision making (DDD) refers to a practice of basing the decisions on the analysis of data rather than purely intuition (Provost and Fawcett, 2013). This is not an all-or-nothing practice; firms differ in their practice of DDD to greater or less degree (ibid). In a study conducted on measuring the strength of Data driven decision making on a firm’s performance by Erik Brynjolfsson from MIT, concluded that DDD is associated with higher productivity, market value and profitability (Brynjolfsson, n.d.). The data-based decision making is enabled through the employment of data analytics (Sun, 2018). Data analytics or Predictive analytics uses statistical techniques, data mining, and machine learning (Lechevalier, Narayanan and Rachuri, 2014). Predictive analytics can be used for failure prediction, forecasting product demand, cost modelling etc. (ibid). Classic statistical techniques – linear and logistic regression, are still the workhouse of most of the predictive models today (Eckerson, 2007). As discussed, for MTS industries, ML models which are based on ‘regression’ have been suggested for improved production planning (Lingitz et al., 2018). Machine learning uses data mining techniques that makes a machine to learn on its own and predict future (Halsey, 2017). While, the data mining is all about statistics and other programming methods to find the hidden pattern in the available data so that a phenomenon can be explained (ibid). Also, Professor David Lowe of Aston University in his interview had quoted, ‘The politicians diverting millions in AI might not know it, but deep learning models are subsets of statistical semi-parametric inference models’ (Significance Magazine, 2019). One study compared regression with ML and found that regression performed reasonably well against ML (Yildiz, Bilbao and Sproul, 2017).
2.8 Regression Analysis

Regression analysis is a mathematical way of sorting out of required variables that have an impact (Gallo, 2015). They answer the questions such as which factors matter the most, which can be ignored, how do factors interact with each other, and most importantly, how confident are we about those factors (Gallo, 2015). These factors are called variables. The dependent variable is the main factor that is being tried to understand or predict and the independent variables are the factors that are evaluated if there is an impact on the main factor (ibid). In other words, regression analysis establishes a relationship between two or more variables. In a regression model, the expected output value is a function of one or more input variables known as independent variables or predictors or regressors and the output value is known as response or dependent variable (Hocking, 2013). Regression analysis produces a straight line that runs in common through all the data points in a graph between dependent and independent variable, thus producing a solution that commonly fits all the points but obviously with some errors. As Benston (1966) puts, “regression analysis essentially consists of estimating mathematically the average relationship between dependent variable and the independent variable”. That straight line is the best explanation of relationship between the dependent and independent variable (Gallo, 2015). Regression analysis is employed in various sectors to predict the output approximately on which several decisions are made. As Gallo (2015) said, many companies have already employed regression analysis to understand a phenomenon that will improve their business, predict future which can include sales, production hours, human resources etc. and even in decision making. And, amongst the quantitative methods, statistical methods have been shown to prove quite useful in similar approaches, specifically regression (Yildiz, Bilbao and Sproul, 2017).

2.8.1 Multiple Linear Regression Analysis

Multiple regression analysis presumes a linear relationship between the variables (Benston, 1966). In multiple regression analysis, a single dependent variable is modelled as a linear function of multiple independent variables with corresponding coefficients and a constant term (Statistics Solutions, 2014). Hence the mathematical form becomes

\[ y = b_1x_1 + b_2x_2 + b_3x_3 + \ldots \ldots . + b_nx_n + c \] (Statistics Solutions, 2014) …. (equation 1)

where b’s are constant coefficients that express the marginal contribution of each x’s to the dependent variable y and c is the sum of unspecified factors or the disturbances that are assumed
to randomly distributed and assumed as a constant (Benston, 1966). In real world, the independent variables are never always a perfect predictor of the dependent variable and hence the regression line always carries an error term, ‘c’ (Gallo, 2015).

There are different methods that are used to carry out a regression analysis, the method of least squares being one of them. In order to understand this method, it is important to understand what residuals are. In regression, the deviations between the readings from the model and the actual readings are called residuals (Field, 2009). Therefore, a squared sum of these residuals can tell how well a model fits a set of data (ibid). The method of least squares finds a ‘fit’ that has the least sum of residual squares and hence, the best overall ‘fit’. In addition, the residuals are squared so that the positive and the negative differences in the readings don’t cancel each other out (ibid). However, one can face collinearity issues in a regression model. Multicollinearity is when there is a strong correlation that exists between two independent variables. A strong correlation between two or more independent variables can make it hard to find a good ‘fit’ as two variables are perfectly correlated, there can be many solutions that will work equally well (ibid). In addition, it is seldom that in real life, a data set will conform to having no collinearity at all.

However, a statistical technique known as ‘Principal Component Analysis’ can help in removing the multicollinearity. PCA can be used to simplify a data set (Wold, Esbensen and Geladi, 1987). In Principal Component Analysis, the existing independent variables are transformed into principal components. These principal components are independent of each other (Liu et al., 2003). In addition, the components are constructed in such a way that they have a linear relationship with the independent variables. Running a regression on these principal components is called ‘Principal Components Regression’ (PCR) (ibid). For example, a data set involving three independent variables will result in three principal components. These principal components can account for different levels of variance in the data set (ibid). Therefore, depending on how many principal components are chosen subsequently for regression, there can be a loss of data in the process.

### 2.8.2 Data in Regression

Regression has been discussed as a power technique that can be used to accomplish multiple things. However, in today’s world, it is not just about the analysis but the data itself. As Corrales, Corrales and Ledezma (2018) pointed out, the availability of data has drastically changed from “scarce to superabundant”. The technological advances have led to an increased
generation of data that until now, was not available. In fact, it is estimated that 1.7 Megabytes of data is generated every second for every person in this world (ibid). Hence, due to this increased generation and availability of data, ‘data pre-processing’ has taken a more important role. This pre-processing mainly consists of ‘cleaning up’ the raw data (Corrales, Corrales and Ledezma, 2018). The following data quality issues are discussed that are associated with regression models in general (ibid):

1) Missing Values: When one or more variables do not contain any value. This can be due to faulty data collection methods. This problem usually occurs at the source

2) Outliers: An observation in the dataset is usually considered an outlier if it deviates too much from the rest of the dataset.

3) High dimensionality: This is usually referred to when there is a high number of variables in the dataset. This can affect the quality of the final result and subsequently, decrease the performance of a regression model.

4) Redundancy: This problem refers to the instances in the dataset when there is repetition of the data. This can again have a detrimental effect on the quality of the data.

Corrales, Corrales and Ledezma (2018) discussed the issues and classified it as ‘noise’ in a regression model. The study also suggests different strategies to deal with the ‘noise’ in order to improve the quality of a regression model. For ‘Missing Values’, ‘Imputation’ is suggested as a counter-approach. This can involve the deletion of certain data values if the missing values are not found. However, there is also the possibility of finding an appropriate value for the missing data from the dataset itself depending on the conditions. In addition, the missing variable can also be treated as a dependent variable while running a regression to cope with the absence of data (ibid). An outlier detection algorithm is suggested to find the outliers in dataset from one of the numerous algorithms available. In addition, regarding the redundancy in a dataset for regression, removing the redundant values in a dataset is suggested in order to increase the quality of the model. When it comes to high dimensionality, dimensionality reduction is recommended to find the most useful attributes in a regression model that can represent the dataset. This can lead to higher accuracy in the model (ibid). The ‘Principal Component Analysis’ that was discussed previously in this section is one of the techniques which can be used for dimensionality reduction as it projects the existing variables or attributes to an orthogonal space and thus, can assist in reducing the number of variables in a regression.
model (ibid). All the techniques have been found to greatly assist in dealing with data quality in regression models and subsequently assist in better regression models.

2.8.3 Output Analysis in Regression

There are a lot of inferences that can be made from a regression analysis output such as descriptive statistics, significance of the independent variables, plots of forecast and residuals, out-of-sample validation (Nau, 2019). This study will mainly focus on the following outputs as they are the most relevant for answering the research questions.

1. R-Square values

2. Values of the estimated coefficients

3. Variance Inflation Factor (VIF)

*R-Square*

Assessing a ‘goodness of fit’ is an important process in using regression models as it describes how ‘well’ the model fits the data (Field, 2009). R-square is a statistical measure that indicates how close the resulted model replicates the data points. In other words, how close is the fitted regression lines to the plotted data points (ibid). In general, higher the R-square, better the model fits the data.

*Values of the estimated coefficients*

The ‘predictor’ variables, which are the independent variables in a regression equation, are always accompanied by a coefficient in a regression analysis (Field, 2009). The value of these coefficients determines how the dependent variable will change, if there is a change in the independent variable which the coefficient accompanies (ibid). For example, in equation 1, the value of \( b_1 \) will subsequently explain how the dependent variable \( y \) changes owing to a change in \( x_1 \). Subsequently, if a coefficient of an independent variable is zero, it might mean that the independent variable in question does not have an effect on the dependent variable at all (Field, 2009). Since these coefficients explain the relationship between the dependent and the independent variable, it can be observed that the sign of the coefficients indicates the correlation between the variables. If the coefficients turn out to be positive, then, there is a positive correlation between the independent variable and the dependent variable, while, if the coefficients turn out to be negative, then, there is a negative correlation between the two.
Variance Inflation Factor

The Variance Inflation Factor or VIF is a diagnostic measure that discusses the collinearity between the independent variables (Field, 2009). As explained before, collinearity between independent variables in a regression analysis can be an issue as if the collinearity is high, then there can be multiple solutions or multiple values of the associated coefficients that can provide an equally good ‘fit’. When such a correlation exists within the independent variables they are not ‘independent’. Therefore, an increase in collinearity pose problems in obtaining accurate values of coefficients (ibid). However, it has been noted that perfect non-collinearity is practically unattainable. VIF values starts from 1 and doesn’t have the higher end. VIF of 1 indicates that the independent variables are not correlated and hence no multicollinearity exists. As a rule of thumb, VIF of less than 10 is acceptable and more than 10 indicates high multicollinearity (Heckman, 2015).

2.9 Statistical Software for Regression Analysis

There are many statistical softwares that could perform a regression analysis. MiniTab is a well-known statistical software for performing advanced statistical analysis as well as for implementing quality improvement methodologies (Alin, 2010). This software had received much recognition and awards too (ibid). MiniTab has a clean and very interactive graphical user interface (GUI). Many leading industries have employed MiniTab, and authors like Cintas, Almagro and Martorell Llabres (2012), Kenett, Zacks and Amberti (2014) had books on Industrial statistics with MiniTab. R is a high level language which is recognized as one of the most powerful and flexible statistical software environments and is now rapidly becoming a benchmark for quantitative analysis and statistics (Crawley, 2007). R can provide applications ranging from simple regression to multivariate analysis (ibid).

2.10 Employee Acceptance

Since the ETO industry is heavily reliant on manual labour and manually conducted operations, it can be said that the employees must play a big role in an ETO organization. Hence, it must be considered important to include the employees in any major organizational change in an ETO company, rightfully so. In fact, the lack of dialogue and communication between the employees of an organization and the organization itself has been deemed as one of the major obstacles in reluctance towards change in an organization (Anton, Camarero and San Jose, 2014). In addition, the same study found that when a new technological change was implemented in an organization, even when there was not much reluctance on behalf of the
employees, no major benefits were reaped from the change owing to the lack of support from the employees (ibid). Hence, it is evident that the support of the employees while implementing any change is an extremely important aspect. However, a severe lack of employee acceptance in the context of the ETO environment was noted in the literature.

2.11 Standardization Programs in ETO

As has been mentioned before, ETO companies deal with extremely complex products that are built based on the specifications of the customer (Vollmar and Gepp, 2015). In fact, it has been noted that ETO companies have been trying to deal with complexity, while also struggling with profitability. One of the strategies that has helped other industries deal with this increasing complexity of operations while also increasing profitability has been standardization (ibid). Despite finding their origins in the product business, standardization programs have been gaining popularity in the ETO business as well as they are seen as a possible way to find a solution to this problem. However, as the study by Vollmar and Gepp (2015) noted, despite finding an increasing interest, there has been a deficit in terms of theoretical knowledge about the methodological aspects of standardisation in the context of an ETO company. The study developed a possible ‘base’ framework that can be built upon for a possible standardization program in the context of the ETO industry (ibid).
3 METHODOLOGY

This section discusses the methodology followed during the course of this study. It begins with discussing the research methodology and subsequently explains the research design and research approach that was followed for this study. The data-collection and analysis methodologies have also been discussed. It is followed by a discussion on the generalizability, reliability, replicability, validity as well as ethical implications of this study.

3.1 Research Methodology

This research followed a mixed method approach. It required a combination of qualitative and quantitative approaches along with a literature study. According to Shorten and Smith (2017), a mixed method approach allows the collection and analysis of both qualitative and quantitative data and also allows to explore wide perspectives of the research questions. Within mixed method approach, this research took explanatory sequential design approach. According to Plano and Clark (2011) as cited in Subedi (2016, p. 3), an explanatory sequential design begins with quantitative data collection and later qualitative data is collected to elaborate on quantitative results. The primary reason for this research method is to provide general results from quantitative analysis and later using qualitative data to refine, extend or elaborate the quantitative results (Subedi, 2016). The process chart referred from Subedi (2016) can be referred to in figure 1.

![Figure 1: Explanatory Sequential Design](image)

3.1.1 Quantitative Methodologies

Initially, a quantitative approach was employed for data collection and analysis. Quantitative approach is a method of research that depends on measuring variables using a numerical system and analysing the obtained measurements using any of the statistical models and interpreting the relationship and correlations within the variables itself (Lucas-Alfieri, 2015). This data was
collected primarily through detailed production plans and detailed engineering plans which was later analysed using statistical softwares.

3.1.2 Qualitative Methodologies

Qualitative research produces “descriptive data such as observations of behaviour or personal accounts of experience” (Lucas-Alfieri, 2015). Qualitative data was collected through in person and virtual interviews. Since RQ2 was explorative in nature, it was important to give the opportunity to the interviewees to answer the questions freely (Bryman and Bell, 2011). Hence, a semi-structured interview with the production planner along with several unstructured interviews were conducted with production manager, production engineer, winding shop manager, and shop floor operator. The results from the interviews were used to answer RQ2 and helped to refine the results from statistical analysis to answer RQ1. The data collection process will be explained in detail subsequently in section 3.4.

3.2 Research Design

With any academic research, a researcher is usually posed with two options, one can either choose to study a lot of cases superficially, or study one case intensively (Gerring, 2006). With the information presented in previous sections, it is evident that this study relied on a wide variety of data to analyse and hence, answer the research questions. Bryman and Bell (2011) have shown the importance of feasibility in a study and how it is extremely important to understand and limit the scope of a study in order to produce quality results. Gerring (2006) describes it best by discussing how sometimes, in-depth knowledge about one case can be more helpful in understanding a phenomenon as compared to superficial knowledge about a lot of cases. For this study in particular, in addition to the aforementioned reasons, another important aspect to consider was the availability of data. As will be explained subsequently in section 3.4 and 3.5, this study depended immensely on organizational data such as production plans as well as the ability to conduct interviews with the employees of the company. Ensuring such high levels of access to data in different ETO companies can be extremely difficult to acquire. Hence, to answer the research questions, a single-case study research design was employed. According to Bryman and Bell (2011), a case study approach is an in-depth analysis of a single case and the term ‘case’ is frequently associated with workplace or organization. This research investigated the effectiveness of using regression analysis for man hour estimations which is one of the primary factors of production planning in ETO industries as such companies rely on
manual labour extensively. This research was conducted at the winding shop of an ETO company where the primary activity is to produce the windings of a transformer.

RQ1 revolving around regression model was answered with the help of database created in MS Excel that primarily consisted of historic data (past projects executed) by the case company and analysis was carried out in MiniTab and R. Through statistical software, R square values, value of the coefficients and Variance Inflation Factor (VIF) were studied and analysed, as they were significant to build the regression model accurately. While RQ2 explored the factors that could have an effect on whether or not, such models are effectively adopted in the ETO industry. The empirical data for it was mainly gathered through interviews and brainstorming.

<table>
<thead>
<tr>
<th>RQ</th>
<th>Literature Review</th>
<th>Interview</th>
<th>Organizational Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ 1</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>RQ 2</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

*Table 1: Data sources used to answer the research questions*

### 3.3 Research Approach

Ontology is the study of being and deals with the structure of the reality as such (Crotty, 1998). And Al-Saadi (2014), discusses, ontology concerns an individual’s beliefs about the kind and the nature of reality. While, epistemology is a way to look at the world and make sense out of it (Crotty, 1998). Epistemology deals with the nature of knowledge along with its possibility, scope and legitimacy (ibid). And, Al-Saadi (2014) discusses, epistemology deals with the assumptions made to understand the world.

Ontology and epistemology take two paradigms each, Objectivism and Constructionism; and Positivism and Interpretivism respectively. According to positivism and objectivism, the truth is always static and is always objective (Al-Saadi, 2014). Hence, the result of the research is independent of the researcher(s). And, according to this view, experiences, empirical knowledge and knowledge derived through an individual’s senses are genuinely regarded as knowledge (Al-Saadi, 2014). To counterpart this view, interpretivism and constructionism, regards perceptions and interpretations over observations to understand a phenomenon (Al-Saadi, 2014). Hence, the research results will depend on researcher(s) assumptions and perspective. Thus, going by the definition, this research will take the path of positivism and objectivism as this research uses results from softwares which will be independent of the
researcher’s perspective on the issue and also the results were refined only based on the opinions of expertise related to the study which is independent of the researcher(s).

A part of this research relied on existing literature on the ETO environment and some of the idiosyncrasies that were associated with an ETO environment. RQ1, for example, relied on the literature that discussed the relationship between man-hour estimations and production planning combined with the regression-based approaches that had been taken in the mass-production environment. A deductive research approach is usually associated with using existing literature and theory for testing purposes instead of generating new theoretical knowledge (Kovács and Spens, 2005). However, this research also focused on aspects that had not been discussed in the literature. Noting a severe lack of production planning methodologies in the ETO environment as well as the absence of using regression as a basis for man-hour estimation in the ETO environment instead of instincts, this research investigated this issue with the help of the research questions. Therefore, it was not purely deductive. With regards to the inductive approach, according to Bryman and Bell (2011), a theory is the outcome of an inductive research where generalized inferences can be drawn out of observations. And according to Kovács and Spens (2005), an inductive approach begins with empirical observations prior to any theoretical framework. This research investigated the usage of regression-based techniques to estimate man-hours in an ETO environment that would subsequently assist in developing production planning frameworks for this environment. Hence, there are inductive elements to this research as well. But this research cannot be classified as inductive either since existing theoretical concepts were being used in order to dive further into the research questions.

Therefore, this research can neither be classified as purely deductive nor purely inductive. Hence, an abductive approach was used since this research took advantage of both, the literature as well as empirics. In fact, it has been pointed out that abductive approaches, especially in case studies, have the probability of yielding more benefit as compared to other approaches (Dubois and Gadde, 2002). This is because it not only uses the existing theoretical models, but also focuses on the ‘empirical world’ (ibid). In addition, abductive approaches have also been shown to make a significant contribution to scientific knowledge (Awuzie and McDermott, 2017).

3.4 Data Collection

Unfortunately, there was not enough relevant literature regarding a structured approach for data collection in ETO to estimate man-hours correctly. However, there had been a study in the ship-
building industry that discussed similar issues regarding the lack of data-driven decision-making in man-hour estimations (Wan Abd Rahman, Mhd Zaki and Abu Husain, 2019). By using a combination of qualitative and quantitative means, the study was able to improve the accuracy of man-hour estimations (ibid). Hence, a similar approach was deemed beneficial for data collection for this study. Planning had been identified as one of the most complex business processes, especially production planning in manufacturing industries. Therefore, for an effective planning process, a diverse database of information is required (Mauergauz, 2016). For production planning, archives of data which are associated with production as well as the design of the product being manufactured/assembled are a basic requirement (ibid). Therefore, it was beneficial to diversify the data sources as much as possible.

This study relied heavily on data collected from the several sources. These data sources could be classified as primary and secondary data sources. The data from the secondary sources were collected and compiled to build the regression model for man-hour estimations. This data was quantitative in nature. However, several unstructured interviews and a semi-structured interview were taken during the process so as to aid the secondary data. The data that were gathered from the interviews were treated as primary data. It is beneficial, however, to point out the reason for non-reliance on primary data sources for the regression model. The secondary data sources included detailed production plans and engineering plans. The data was compiled in a MS Excel file after noting down the total production time for a winding from detailed production plans. This was the ‘dependent’ variable in the analysis. In addition to the total production times, to build the regression model, it was also important to include the several operations which would be included in the regression analysis as independent variables. The values of these independent variables were derived from the detailed engineering plans. These data sources were classified as secondary source since it was provided by the case company. This type of data was used extensively in the study instead of observational studies owing to the extremely long lead times.

In addition, for a regression analysis, there is a statistical requirement in every statistical software. There should be at least one more dataset than the number of variables. So, for example, if there are 10 variables in a regression equation, there should be at least 11 datasets in the data matrix. A regression model with a total of 26 variables was constructed for this study. Therefore, a minimum of 27 data entries would be required for this study, which was not suitable considering the time constraint. Hence, historical production data was used which was based on the projects that were carried out between 2017 and 2019. To add to it, ETO is an
industry that dealt with a lot of complexities, hence adding more data would aid in identifying the pattern as well as yield accurate results.

![Diagram of data sources](image)

*Figure 2: Various types of data sources employed for this study*

### 3.4.1 Literature Review

A literature review was performed in order to get a deeper understanding of the existing scenario regarding the man-hour estimations in the ETO industry. Bryman and Bell (2011) discussed the importance of a critical literature review. One of the important points that had been discussed is how a literature review can be a means of showing the importance of the research questions (ibid). A literature reviewing methodology highlighted by Bryman and Bell (2011) was followed that assisted in conducting a detailed and unbiased review of the existing information. Literature review of this study began with explaining the ETO industries and the complexities that were naturally associated with such industries. ETO industry was compared with mass manufacturing industries to emphasize on the complexities. Then, production planning and its requirement in an ETO industry was explained in detail. Followed by, the process of bidding and man-hours estimations and why they were significant for production planning were explained. After discussing the persistent issues in the ETO, it was possible to show that there was a severe lack of data-driven decision-making in the ETO industry because of the complexities associated with it. Then, a review was done on how mass manufacturers dealt with predictive analytics which basically means estimating any variable or a factor for future projects. After concluding that regression analysis forms a backbone for most of the
advanced technologies in today’s world, a review was made on regression analysis in order to brainstorm how it could be employed in ETO industry. Finally, two major strategies, the importance of employee acceptance in any change management and process standardization, that could help ETO with profitability and growth were discussed. In addition to highlighting the problem, the literature review also assisted in understanding the methodologies such as regression that was used in this study. Uppsala University’s access to different databases and Google Scholar were used to look for the relevant articles. Keywords such as ETO, Engineering to Order, Production Planning, Man hour Estimations, Regression analysis etc. were used. The literature review assisted in understanding the contextual information as well as provided a stepping-stone to further investigate the empirics and answer the research questions accordingly.

3.4.2 Quantitative Data Collection

Bryman and Bell (2011) pointed out how organizational documents can help in understanding the managerial decisions that had been taken in the past in a case study research design. Therefore, in this research, organizational documents were of utmost importance. Data in terms of production plans and the design specifications were one of the major sources of data for this study. This data was collected, as well as analysed in a quantitative manner. As pointed out already, production planning is complicated and production-related documents can be of great assistance in such a case. In addition, there was a need for several documents related to the latest finished projects, what products were assembled during these projects, the time it took, the design of the products so as to understand the different operations better. Some documents such as the annual report, press releases etc. are easily available in the public domain, whereas other documents such as production schedules in this case, are not (Bryman and Bell, 2011). However, access to these documents was assured for this research. The procedure that was used to analyse the quantitative data has been explained in detail in section 3.5.

3.4.3 Qualitative Data Collection

As pointed out before, for this study, it is best to diversify the data sources as much as possible. Since there was a lot of reliance on secondary data sources such as production plans as well as engineering plans, there was a need to make sure that the data and its interpretation was accurate. In certain cases, in order to avoid confusion and inaccurate data entries, unstructured interviews were taken with the production planner at the case company. The unstructured interviews also complemented in improving the statistical results to answer RQ1.
However, for RQ2, there was a need to understand the context of the ETO environment in-depth. Interviews have become one of the more common and widely used methods of data collection. Interviews help in understanding how people think and perceive a phenomenon. Unlike the quantitative ways of collecting data, interviews help in examining a process in-depth (Frances, Coughlan and Cronin, 2009). For this purpose, semi-structured and unstructured interviews were used over structured interviews. According to Bryman and Bell (2011), semi-structured and unstructured interviews cause less hinderance to the interviewee in the process of answering the questions which subsequently results in getting more in-depth results from the interview. In addition, a lot of decisions in the ETO industry are based on human intuition. Therefore, it can be very tough in understanding the logic behind the various decisions that take place if a structured interview was conducted. Hence, unstructured and semi-structured interviews helped in coping with these circumstances.

In addition, RQ2 dealt with understanding the factors for successful adoption of DDD models in the ETO industry. And in order to understand what these obstacles might be from an empirical standpoint; it was important to know how people working in the industry perceive the problems to be. However, it was not possible to interview everyone associated with the ETO industry to understand the complications. This is where sampling came in. Sampling enables in selecting a representative portion of the population for the study (Bryman and Bell, 2011). Therefore, in this study, instead of studying the whole ETO industry, sampling assisted in selecting a representative portion from the case company. Purposive sampling technique was used as it assisted in selecting a sample that was relevant to the study. Purposive sampling can help in selecting the unit of analysis that is relevant to the research (Bryman and Bell, 2011). In this case, for example, interviewing a safety specialist might not be useful when understanding production planning procedures in an ETO company as compared to a production planner. Unfortunately, when purposive sampling is conducted, the results of a study are not generalizable to a population (Bryman and Bell, 2011). However, the insights gained using this study could still contribute to the scientific knowledge database as not enough is known regarding this environment. Many interviews were conducted, most of them were unstructured and were used primarily to aid the quantitative data. One semi-structured and several unstructured interviews were conducted specifically for the second research question in order to understand the planning procedure in an ETO environment and the factors that could affect successful adoption of the regression model. One of the primary requirements for conducting a semi-structured or a structured interview is an interview guide (Bryman and Bell,
2011), which is not possible or tends to be less effective if constructed with just one or two specific questions. Hence, such conversations which helped this study to get a deeper understanding were treated as unstructured interviews. Unstructured interviews were most of the times, a short call with interviewees or a mail with a few specific questions that sometimes continued as a thread. Hence, the number of such conversations were not specifically counted and as cited in (Bryman and Bell, 2011, p. 467) by Burgees, unstructured interviews have very similar characteristics to a conversation. The details regarding the total number of interviews are presented in table 2.

<table>
<thead>
<tr>
<th>Interviewee</th>
<th>Designation</th>
<th>Number of Interviews conducted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person 1</td>
<td>Production Development Manager</td>
<td>Several Unstructured</td>
</tr>
<tr>
<td>Person 2</td>
<td>Production Engineer</td>
<td>Several Unstructured</td>
</tr>
<tr>
<td>Person 3</td>
<td>Production Planner</td>
<td>1 Semi-Structured and Several Unstructured</td>
</tr>
<tr>
<td>Person 4</td>
<td>Winding Shop manager</td>
<td>1 Unstructured</td>
</tr>
<tr>
<td>Person 5</td>
<td>Operator</td>
<td>1 Unstructured</td>
</tr>
</tbody>
</table>

*Table 2: Designation of interviewee and number of interviews conducted with every interviewee*

### 3.5 Data Analysis

This section of the thesis describes how the collected data was modified and used for analysis. It is divided into three steps; the first step explains about construction of a database for conducting the regression analysis where data cleaning techniques were used to modify the collected data to produce accurate results. The second step is about conducting the regression analysis and producing results. The final step explains about analysis and evaluation of results to answer the research question.

A database was made in MS Excel based on the data collected from the secondary data sources. Before conducting the regression analysis on the data collected from detailed production plans and detailed engineering plans, data cleansing was carried out. As discussed in the literature review (Section 2.8.2), various techniques suggested by Corrales, Corrales and Ledezma (2018) were followed to modify the input data in order to produce accurate estimates.
Quantitative analysis was carried out extensively in MiniTab statistical software while R was also used to analyse a regression method which was not possible in MiniTab. Most of the analysis were done in MiniTab where linear regression option was used extensively for the analysis. The dependent variable for the analysis was the total production time for producing a winding and the independent variables were the operations that were performed to complete the production. Each operation performed will have a linear effect on the total production time. As the relationship between dependent and the independent variables was linear, linear regression option was used for analysis. In R, “faraway” library package was used to call a function called NNLS (Non-Negative Least Square) method. This was used to force the software to deliver positive values to the coefficients as the coefficients cannot be negatively correlated with production time, because, if the coefficients were negatively correlated, it implies, if a specific operation was carried out, the total production time would decrease which could never be the case.

While for the qualitative analysis, an interview guide for a semi-structured interview was prepared with casual questions as well as main study questions so that the interviewee had more time to share their experience. Several unstructured interviews were also conducted to better understand the current production planning scenario in the ETO industry and various factors that can affect a successful adoption of models such as the one mentioned in this study. The unstructured interviews also ‘aided’ the quantitative data. Detailed explanation will follow in section 4.11.

3.6 Generalizability of this Research

“To generalize is to claim that what is the case in one place or time, will be so elsewhere or in another time”(Payne and Williams, 2005). Though, the research is about ETO environment, the research cannot be generalized to all companies in ETO category as the variations vary between companies, as pointed out by Adrodegari et al. (2015). As cited in Bryman and Bell (2011, p. 61) by Lee, Collier and Cullen (2007), particularization constitutes the main strength of a case study research than generalization. The goal must therefore be to focus on the uniqueness of the case and develop a deep understanding of the complexity (Bryman and Bell, 2011). Confidentiality also plays a major role in generalization. The company’s name, products, production parameters and the extent to which each parameter differ are confidential which makes generalizing difficult to other companies even if they fall under same classification as the case company. In addition, purposive sampling has been used for qualitative data collection.
Purposive sampling weakens the generalizability of a study (Bryman and Bell, 2011). However, it should be noted that despite the weak generalizability, this study still offers some key insights into the ETO environment. This study should be looked from the perspective of a steppingstone instead of trying to accomplish an end result.

### 3.7 Reliability and Replicability

Reliability of a study is concerned whether the results from study are repeatable (Bryman and Bell, 2011). Reliability is quite essential with the quantitative research (ibid). While, for a qualitative approach, the results of the study would depend on the questions and the people being interviewed or surveyed. Since, a combination of qualitative and quantitative approach was followed for this study, the reliability is moderate.

Replicability of a study is concerned with the degree to which the results of a study are replicable if they are conducted again (Bryman and Bell, 2011). Since a combination of interviews and statistical softwares were used to answer the research questions, the replicability of final analysis and results is moderate while the replicability of the results from the statistical software is high with a bit of small variation.

### 3.8 Validity

According to Bryman and Bell (2011), internal validity of a research is concerned with causal relationship between two variables. If the conclusion is x causes y, is it sure that only x is responsible for variations in y and not z. Though production time is the significant output of independent variables considered for this research, internal validity of this research is weak. This is because the case company relies heavily on manual labour. Hence, it is extremely hard to incorporate every single variable that can influence the production time. For example, a newer employee might take more time to carry out a certain operation than a relatively experienced employee. External validity, on the other hand, is concerned with the generalizability of the research (Bryman and Bell, 2011). Hence, the external validity is also weak for this study. However, the analysis methods might still prove useful, provided the idiosyncrasies of the company in question are accounted for.

### 3.9 Ethics

Ethics play a major role not only in day-to-day life, but also in academic projects (Bryman and Bell, 2011). In fact, the Swedish Research Council insists that every student degree project
should meet a set of ethical criterion which include but are not limited to consent and confidentiality requirement (Blomkvist and Hallin, 2015). However, ethics is not merely about fulfilling requirements but also thinking critically about the effects an academic study can cause, intentionally or unintentionally. Awareness & Responsibility constitute a major part of ethical behaviour (Lennerfors, 2019). This study, for example, deals with the usage of historical data for production planning in an ETO environment. As discussed in the literature review, the man-hour estimations that was generated will be subsequently used for numerous applications including price bidding as well as production planning for efficient use of resources. Therefore, it was important to be aware of these use cases and the problems they pose. For example, if the output from the study turned out to be non-reliable, it might pose a problem during the bidding stage which might end up in the case company incurring a loss. More importantly, the estimations will also be used for production plans and schedules. If the time estimates were too conservative or completely wrong, it puts unfair pressure on the people involved in the manufacturing process to cope up with numbers that are unreliable or in some extreme cases, even unrealistic. In this case, whatever the ‘consequences’ might be, there is a ‘responsibility to’ be aware of these conditions (Lennerfors, 2019). In addition, the confidentiality agreement plays a major role. This study extensively used historic data which included the product type as well as its design specifications. These were sensitive data and was used responsibly. In addition, during the research, no harm was caused to any participants. Before conducting interviews, the reason for the interviews were communicated and interviews were conducted only after their consents were received. This research was conducted from an overt perspective.

In addition, another interesting perspective on the ethical implications for this study are based on societal implications, specifically the job market in Sweden and its relation with globalization. A study published in 2013 focused on differences in the quantity of exports and imports in a country, its relation with domestic and foreign labour and its subsequent effects on foreign labour demand (Jiang and Milberg, 2013). There were numerous interesting aspects about this study, but perhaps the most interesting aspect was where the different countries stood in terms of their domestic labour and foreign labour demand. Domestic labour demand in this context meant how much of value creation took place inside the country whereas foreign labour demand meant how much of the value creation took place outside of the country and hence the word, ‘foreign labour’. In the rankings in terms of the difference between the difference in the demand of domestic labour and foreign labour, China, India, Indonesia and a lot of other developing countries were amongst the highest in the ranking. Which means that these countries
are producing a considerable amount of work with the help of their domestic labour. Sweden ranked amongst the last few countries in this study, which means that Sweden is producing more value with the help of foreign labour than domestic labour (ibid). Whether it is good for the economy or bad for the economy is a different area to explore, one can argue that this can result in fewer jobs ‘inside’ the country. This might possibly be because of high labour costs as all the last few ranking countries were other developed countries such as the US, Germany etc. Hence, if this degree project results in a methodology that derives better efficiency in the case company, it could also subsequently result in the production taking place inside the country and hence, creating more opportunities.
4 EMPIRICAL DATA

This section will describe the empirical data that was collected during the course of this study. A basic description of the manufacturing plant as well as the products manufactured in this plant has been discussed. In addition, the process that was followed throughout the study to investigate the possibility of using regression has been described in detail.

4.1 Plant Description

This research was conducted at a multinational company. The company has several divisions which manufactures a range of electrical, electronic and robotic products. This research was conducted in a transformer manufacturing plant located in Sweden. It focusses on winding department of the plant. The plant could be classified as an Engineering-to-Order class, as the company is customer specific in designing and manufacturing their products. In an ETO company, one of the major reasons for planning the production is not only for resource planning but also for submitting the price quotation for the project during the bidding stage. The output production hours from the production planning tool will be used to calculate the man-hours and labour cost for the project, which will eventually be used for the final project quote. The major element in a transformer is the winding which houses complex specifications and designs which can be performed effectively only by the skilful labour and not by any other automatic or semi-automatic machines, as confirmed by the company. But the management was finding it difficult to project approximate production hours due to the complexities and differences involved in the products as well as the differences between the operators themselves. The company had experimented in the past with using regression as a production planning tool wherein the production data was collected from 2015 to 2017 and the regression model was constructed. With this step, some improvements in the production planning process were observed. Despite the model not being as robust, the accuracy of man-hour estimations increased which subsequently trickled down to numerous operations that are carried out on the winding shop-floor. Hence, there was a desire to conduct a more structured and academic study to investigate the possibility of employing regression to predict man-hours, that could reduce the errors during production planning as the decisions are based on data rather than expert’s instincts or experience. This case study will explore the opportunity of using regression analysis, one of the statistical approaches, as a production planning tool in an ETO company by inputting the data of the company’s historic projects. This study will also explore on how regression model could be successfully adopted in ETO industries.
### Products Manufactured

<table>
<thead>
<tr>
<th>Products Manufactured</th>
<th>Heavy duty transformers and other electrical parts</th>
</tr>
</thead>
</table>

### Production Type

<table>
<thead>
<tr>
<th>Production Type</th>
<th>Engineering-to-order</th>
</tr>
</thead>
</table>

### Production Volume

<table>
<thead>
<tr>
<th>Production Volume</th>
<th>Low. Usually project based.</th>
</tr>
</thead>
</table>

### Production Pace

<table>
<thead>
<tr>
<th>Production Pace</th>
<th>Slow</th>
</tr>
</thead>
</table>

### Level of automation

<table>
<thead>
<tr>
<th>Level of automation</th>
<th>Very low. High involvement of manual labour at workstations</th>
</tr>
</thead>
</table>

### Data Reporting and Collection

<table>
<thead>
<tr>
<th>Data Reporting and Collection</th>
<th>Manual. Operators report the data and production planner collects the data</th>
</tr>
</thead>
</table>

### Objective of this study

<table>
<thead>
<tr>
<th>Objective of this study</th>
<th>Performance of a regression-based tool in calculating the man hours</th>
</tr>
</thead>
</table>

**Table 3: Summary of the case company**

#### 4.2 Product Description

The winding division is responsible for the production of the transformer windings. The production processes and setup vary depending upon the type of the winding and specifications. The different types of windings are Helical, Disc and Layered. Every project is divided into subprojects usually classified as A, B, C and D. Each subproject represents a specific winding type in that project.

<table>
<thead>
<tr>
<th>Project Name</th>
<th>Subproject</th>
<th>Winding Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project X</td>
<td>Winding A</td>
<td>Layered</td>
</tr>
<tr>
<td></td>
<td>Winding B</td>
<td>Helical</td>
</tr>
<tr>
<td></td>
<td>Winding C</td>
<td>Disc</td>
</tr>
<tr>
<td></td>
<td>Winding D</td>
<td>Disc</td>
</tr>
<tr>
<td>Project Y</td>
<td>Winding A</td>
<td>Helical</td>
</tr>
<tr>
<td></td>
<td>Winding B</td>
<td>Helical</td>
</tr>
<tr>
<td>Project Z</td>
<td>Winding A1</td>
<td>Layered</td>
</tr>
<tr>
<td></td>
<td>Winding A2</td>
<td>Layered</td>
</tr>
</tbody>
</table>

**Table 4: Classification of projects and sub-projects**

However, it should be noted that Project X, Y and Z do not represent any real case projects, they were used as representations to provide explanations on how the projects and subprojects are classified in the case company. In Project Y, though the winding type is same, the subprojects are different as it will differ in the specifications despite being the same type. However, in Project Z, the windings belong to same subproject (A) and have same engineering specifications. There are still subdivisions within each winding type, which is not a subject of interest of this study. Hence, only a brief overview of the various products would suffice.
4.3 Constructing Database for Regression Analysis

The database construction process that was followed to perform a regression analysis has been described. The data-collection methods that were used have been discussed. The subsequent creation of the database has also been described in detail.

4.3.1 Data-Collection methods

This study primarily dealt with the phenomenon of man-hour estimation in an ETO industry. However, it should be pointed out that the lead times are extremely long. During an interview with a production manager, it was estimated that on a yearly basis, in the winding department, a mere 60-70 products can be produced. However, it was also pointed out that in the last 2-3 years, a lot of effort has been put in to increase the efficiency of the production line and that may or may not have led to an increase in the yearly number of windings produced. So, as discussed earlier, it was a necessity to rely on secondary data sources in order to perform regression effectively. This secondary data sources included the projects that were carried out from 2017 to 2019. Within these projects, various types of windings were produced. There were two secondary data sources, detailed production plans and detailed engineering plans that had been collected from the company.

4.3.2 Detailed Engineering Plans

Whenever a winding was to be constructed, it was fitted onto one of the workstations where different operations were performed until it was ready. These operations were performed a certain number of times depending on the design of the winding. The production planner along with the production department interpreted the drawings of these windings which was supplied by the winding construction (engineering) department. On the basis of the drawing, the interpretations were the sequence of operations to be performed and number of times the operations to be performed. This information was available from the detailed engineering plans that was available in MS Excel files for various projects. Due to confidentiality reasons, original file cannot be shown, while a representation of a similar file is shown in table 5.

<table>
<thead>
<tr>
<th>Operations performed</th>
<th>Number of times operations performed</th>
<th>Estimated Time</th>
</tr>
</thead>
</table>

*Table 5: Representation of Detailed Engineering Plans template for various windings.*
4.3.3 Detailed Production Plans

Detailed production plans were primarily used by the production planner to plan a production of a project. The production plans usually consisted of two sets of data, estimated and actual. Estimated data were usually provided before the start of the production, while actual data was collected after the production was completed. Usually a production planner collects the data from the shop floor. Before starting the production, the estimated parameters - operations that needs to be executed, the shift number for every operation, number of operators and estimated time for each operation, would be filled in by the production planner and would be handed over to the shop floor. The estimated time was derived from the detailed engineering plans. While after the production is completed, the details regarding the different shifts such as total time worked, operations performed, day on which an operation was done, and comments are filled in by the operators. Comments would be written if there were any change from the planned schedule or any unusual incidents were reported. Due to confidentiality reasons, original file cannot be shown, while a representation of similar file is shown in table 6.

<table>
<thead>
<tr>
<th>Operations</th>
<th>Number of Operators</th>
<th>Shift Number</th>
<th>Estimated Time</th>
<th>Actual Time</th>
<th>Day of the operation</th>
<th>Comments</th>
<th>Difference between Estimated and Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>These data were given prior to the production by production planner to the shop floor</td>
<td>These data were filled in by the shop floor operators after completing every operation. These data were collected from the shop floor by the production planner</td>
<td>This analysis was done by the production planners later</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 6: Representation of Detailed Production Plans template for various windings.*

Using this historical data, a preliminary database was constructed with the project name, the winding type, the total production time and all the operations with number of times they were performed. In total, data of more than 250 windings were included in this database.

In this study, the total production time for all the windings was treated as the ‘dependent’ variable as this research tries to predict the production time for projects. Dependent variables are the variables that a research tries to understand or predict (Gallo, 2015). Production time is the total time that is required to properly handle an item in a process and it includes, preparation
time, run time, move time, inspection time and put away time (Langstrand, 2016). Hence, for this research, since the dependent variable is the production time, the independent variables or predictor variables are the operations on the windings. However, it should be noted that the operation times in a manual production line can also be dependent on numerous factors such as the expertise of the person performing the operation. However, in order to keep the scope of this study feasible, it was in best interest to keep the analysis focused on the more significant variables such as the type of operation being performed instead of who is performing it. Hence, the preliminary database template can be seen in table 7.

<table>
<thead>
<tr>
<th>Projects</th>
<th>Winding</th>
<th>Actual Production Time</th>
<th>Various Operations performed in the winding</th>
</tr>
</thead>
<tbody>
<tr>
<td>This column mentioned project name which was derived from Detailed Production plans</td>
<td>This column mentioned winding type which was derived from Detailed Production plans</td>
<td>This column mentioned actual production time at the shop floor. This data was derived from Detailed Production plans</td>
<td>These columns mentioned number of times an operation is performed on the winding which was derived from Detailed Engineering plans</td>
</tr>
</tbody>
</table>

Table 7: Preliminary database template

4.4 Data Quality

Data quality is an extremely important aspect for any modelling procedure. A common principle known as ‘GIGO’ or ‘Garbage in Garbage out’ was followed, which subsequently means that the quality of data that is used as an input to a model can greatly affect the quality of output that one might get (Corrales, Corrales and Ledezma, 2018). The previous section (Section 4.3) explained how the database was constructed. However, it was not a simply a matter of compiling data from different sources into a database and performing a regression analysis on it. There were numerous obstacles and issues that lead to the data being not completely usable which led to an iterative process of ‘cleaning up’ the data using the structured approaches and will be discussed in the section 4.5.

Unavailability of data was one of the obstacles for the database construction. The case company had been recording production plans for all the projects since 2015. However, upon further examination, it was observed that in some cases, there were detailed engineering plans available which contained the information regarding the several operations that were carried out during the manufacturing process, but the production plans were not complete. The data collection
methodology may be the culprit here as the data collection is not automated. As discussed in the literature review, ETO companies can have problems with collecting production data because of high reliance on manual labour owing to the customized products. That is the situation for the case company as well. The data collection was done manually by the production planner from the shop floor. Despite the case company’s best efforts, the unavailability of data was still encountered as a problem in some projects. In addition, there were also problems associated with the inaccuracy of data. In case of engineering plans, negative values were found for some operations which should not be the case as they denote number of times an operation was carried out. There were a lot of variances in the data as well. For example, in some projects which included different units of the same windings, the differences in the time that it took to manufacture them was as high as 50%. These variances can be extremely hard to explain owing to the high dependence on manual operations. There can be numerous other factors contributing to this variance. Even though it is important to examine those variances in order to build an accurate regression model, owing to the time constraint, it was not possible to do that. These variations also lead to problems with validation. As the model itself can’t account for all the variances, it was hard to validate whether the model was performing the way it was intended. A structured ‘data cleaning’ approach was followed to improve the accuracy. It helped in making the database usable. However, it should be noted that it was an iterative process. Multiple approaches to the regression analysis were used which will be described subsequently which led to a ‘back-and-forth’ process between cleaning up of data and trying out different approaches to the regression analysis.
4.5 Data Cleansing

Once a preliminary database was prepared, a structured approach was followed to eliminate the data quality issues as discussed in section 4.3. Structured approach discussed by Corrales, Corrales and Ledezma (2018) as shown in figure 3 was adopted to cleanse the data in the preliminary database to produce more accurate estimates.

![Structured Data-cleaning approach, as explained by Corrales, Corrales and Ledezma (2018)](image)

Figure 3: Structured Data-cleaning approach, as explained by Corrales, Corrales and Ledezma (2018)

4.5.1 Missing Values

The process of imputation has been suggested by Corrales, Corrales and Ledezma (2018) to deal with the missing values. Four approaches to deal with imputing missing values were deletion, hot deck, imputation based on missing attribute and imputation based on non-missing attributes (Corrales, Corrales and Ledezma, 2018). In this research, the deletion methodology was followed to deal with missing values. Deletion is eliminating the whole record if any of the value is missing (Corrales, Corrales and Ledezma, 2018). Deletion was selected among others because hot deck and imputation based on missing attribute dealt with assigning a value to the missing variable (ibid) and estimating a value is very difficult to decide in an ETO industry without much experience and knowledge, and the probability that it can turn out to be an error
was high. In some cases, the actual production time was missing which cannot be assumed based on central tendency measures or some random value from the same dataset. Moreover, this study was based on more than 250 records and the missing data were only around 20 which motivated to choose deletion option as there wouldn’t be a significant change in the results if the missing data were removed.

4.5.2 Outlier Detection

An outlier is a score that is very different from other datapoints (Field, 2009). They are problematic for statistical analysis because they bias the resulting model (ibid). Since, the production time depended on the operator and the circumstances in the shop floor, it might house outliers. For instance, three same windings were completed in x hours, while one other same winding with same specifications was completed in x+500 hours which looks obviously impossible, but certain circumstances such as unavailability of raw materials, machine breakdown, operator’s knowledge etc could be the possible reasons for such an outcome. Thus, in the dataset, possible outliers can be present only in the production time (dependent variable) as the independent variables were from the engineering drawings, there were no possibilities of an outlier. Since, the preliminary database was constructed in MS Excel, Quartile function was used to identify the outliers (Alexander, Kusleika and Walkenbach, 2019). Quartile function provides the better estimates of the actual population and a more accurate view of outliers (ibid). In this study, to identify the outliers, leveraged interquartile range (IQR) method was followed (Alexander, Kusleika and Walkenbach, 2019). An interquartile range is the data that lies at the centre at 50% or in other words in between 25th percentile and 75th percentile (ibid). The term “leveraged” denotes that the centre part is expanded, and the fences are put on both sides and the values outside this fence are outliers (ibid). The outlier detection process began with sorting based on projects. Then, subprojects with same winding type and same engineering parameters were sorted. Sorting in this way was necessary because, if the same winding with same engineering specifications was manufactured for a project, then the production time for all these winding must revolve around the mean value (50th percentile). The data was identified if the value lied above 75th percentile and below 25th percentile. After identifying the outliers, they were removed from the dataset. Since this study followed the structured data-cleansing approach by Corrales, Corrales and Ledezma (2018), removing outliers was a part of this process.
4.5.3 Duplicate Values

The preliminary database didn’t contain any duplicate values as the database was constructed according to each project and winding type. But there were few inaccurate data - negative values in the independent variables. There were no probabilities of carrying the negative values. Hence, to deal with this, the negative value instances were removed from the database, an approach discussed by Corrales, Corrales and Ledezma (2018).

4.5.4 Dimensionality Reduction

Dimensionality reduction reduces the number of independent variables so that the final set of variables represent the dataset accurately without any redundancy (Corrales, Corrales and Ledezma, 2018). Four approaches have been discussed to reduce the dimension which are Filter, Wrapper, Embedded and Projection (ibid). In this study, filter methodology has been adopted to reduce the number of independent variables. Filter method helps in selecting features that are independent of the regression model (Corrales, Corrales and Ledezma, 2018). This methodology was carried out only with approach 5 (Section 4.11).

4.6 Construction of Regression Model

As discussed in Section 2.8, a regression analysis enables the possibility of finding out a relationship that exists between variables (Zou, Tuncali and Silverman, 2003). The variables can be both continuous and categorical (ibid). Since this study had continuous variables, a regression analysis could be performed in order to understand the relationship between the variables. Mathematically, the regression model looked as follows:

\[ Y = \lambda + aA + bB + cC + \ldots + zZ \quad \ldots \ldots \quad \text{(equation 2)} \]

Y= Total production time (Dependent; Input from detailed production plans)

\( \lambda \)= Intercept or Constant value

A (Uppercase)= Parameter of operation A i.e. how many times operation A is performed (Input via detailed engineering plans)

a (Lowercase) = Coefficient of operation A in the model. Describes the time taken (man-hours) for performing operation A every time (Unknown value).
Thus, the product of uppercase and lowercase letters (aA, bB etc) will result in the time taken to complete an operation. The total production time would be the sum output of this model. Now, the value of coefficients as well as the intercept is unknown. However, as will be explained subsequently (Section 4.7), the intercept can be removed from the model and the coefficients can be calculated accordingly. If the intercept is removed, the coefficients in that case would represent the time it takes for the respective operation to be performed once. The reasons for doing so will also be explained. Hence, the historic actual production time is derived from detailed production plans and substituted in place of Y. After the coefficients are found, the total production time can be estimated for future projects. Since the aforementioned equation has many independent variables it can be classified as multiple linear regression model (Field, 2009). Once all the data was collected and cleansed, the final database for regression analysis was completed. To respect the confidentiality, various operations (independent variables) were renamed as alphabets. The output factors from the regression analysis that would be studied for this research are discussed in following sections.

4.6.1 R-Square Values

As discussed in Section 2.8.3, R Square values indicates how close the resulting model represented the input datapoints. Higher the R square value, better the model.

4.6.2 Values of Estimated Coefficients

In this study, the production time was the dependent variable, and the independent variables were the number of times the operations being performed on a winding in a project. Hence, it is obvious that, all the variables must have a positive correlation with the dependent variable as each time an operation was performed it would result in the increase in the production time and vice versa can never be the case. And, zero can never be the coefficient, as it makes the operation (independent variable) to be redundant, while, in actual shop floor that operation is being performed. Hence, non-zero positive coefficients were aimed for all the independent variables with respect to production time (dependent variable).

4.6.3 Variance Inflation Factor (VIF)

As discussed in Section 2.8.3, VIF was used to detect the presence of any multicollinearity in the regression model.
4.7 Approach 1: Ordinary Least Square Method

Recalling from the discussion on the current literature (Section 2.8.1), Ordinary Least Square Method is one of the most common techniques used for a regression analysis. This method tries to find the best ‘fit’ by minimizing the value of square of ‘residuals’, which is the difference between the modelled and the actual value (Field, 2009). The database that had been constructed was used for this analysis. Minitab was used to perform the analysis as it offered an intuitive graphical interface while also offering a lot of different options to modify the model as per the need. After an unstructured interview with the production planner, it was observed that there was a need for a simple implementation of regression as a tool owing to time constraints in the production planning process. This was a great suggestion as ETO industries already face obstacles in production planning and having a complicated production planning tool might make things more complicated. Hence, it was aimed that the end-result of this study should be easily implementable. Referring to general regression equation, there is a constant term (intercept), $\lambda$. Minitab offers the possibility to remove this intercept value from the model and hence, only get an output in terms of coefficient values. This was deemed useful as removing the intercept value eases the possibility of implementing the end result by making the model less complicated. If the intercept value is removed, the planner can simply multiply the coefficients from the output with the number of times the respective operation is performed. This also makes it easier to validate the model since the coefficients in this case would translate in real life to the time it takes to perform the respective operation once. Moreover, it also makes it easier to discuss the coefficients with the production planner as well as the operators and get their opinion on how accurate they deem the coefficients are. The reason why qualitative feedback was still given importance despite the fact that this study was aiming for data-based decision making is because this area is still unexplored territory for academic literature on ETO environments. In addition, as discussed in the ethics section 3.9, this study might have direct consequences on not only the planning procedure for the case company, but also the production operators. Therefore, it was of utmost importance that the experience of production operators is accounted for and is used as complement to the data instead of the data replacing experience completely. This approach was also shown to be more effective in a study that was performed in the ship-building industry (Wan Abd Rahman, Mhd Zaki and Abu Husain, 2019). The results were obtained using Minitab. The output can be found in Appendix A.
AT = -21.3A - 17.77 B + 0.872 C - 1.66 D + 2.839 E + 4.62 F + 1.998 G + 4.88 H + 0.1041 I - 0.411 J + 0.662 K - 3.895 L + 0.043 M + 0.00781 N + 0.03568 O + 0.0569 P - 1.12 Q + 3.16R + 1.171 S - 0.740 T - 1.800 U + 9.7 V + 36.72 W + 17.9 X + 0.391 Y - 28.25 Z (Equation 3)

‘AT’ in the regression equation (Equation 3) is the dependent variable which is the total production time. The regression equation above denotes the relationship between the different operations which had been named as A, B, C, D…Z. Due to a confidentiality agreement, it is not possible to disclose the database as well as what kind of operations these variables denote. It can be observed that some of the coefficients for certain operations are negative. As the intercept has been removed and the output is only in terms of the coefficients, referring back to the regression equation, it can be said that the coefficients denote the average time it takes for the respective operation to be performed once. Therefore, there should be no negative coefficients. Hence, this result was not acceptable merely from a practical standpoint. So, there was a need for another approach. In addition to this, it can also be noted in Appendix A that the VIF values or the ‘Variance Inflation Factor’ that denote the collinearity of the different independent variables is quite high. This can significantly have an effect on the quality of the regression output. However, interestingly, the R square value of the model is quite high, approximately 95%. This translates to the fact that this regression model can account for almost 95% of the variances in the data. Even though this number by itself is not conclusive, it was still interesting to note that it was in-fact possible to account for the variances in the data despite the problems regarding negative coefficients as well as high VIF values.

4.8 Approach 2: Non-Negative Least Square (NNLS)

From the previous analysis method, there were issues of negative coefficients which this study wasn’t aiming for. Hence, to counterpart the issue, NNLS methodology was adopted to construct the regression model. A Non-Negative Least Square (NNLS) regression solves linear least square problems where the resulting coefficients are constrained to be non-negative (Bombara et al., 2009). NNLS finds its applications from control systems to AI (ibid). MiniTab didn’t have the option to conduct NNLS, hence R was used to try this approach. This approach succeeded in producing positive values (at least non-negative) for all the independent variables, while on the other hand, this method resulted in producing zeroes for many independent variables which should not be a case, because, it implies it takes 0 hours to perform an operation. Moreover, the issue of multicollinearity was still persisting. The results of this approach from R is available in Appendix B.
4.9  Approach 3: Principal Components Regression (PCR)

As was observed in the previous two approaches, there was high collinearity between the independent variables. While in discussion with the production operators and planners, it was clarified that the variables with the high VIF values were in-fact independent of each other. Hence, the reason for the high VIF values was hard to understand. Nevertheless, as discussed in the literature review (Section 2.8.3), high VIF values mean high collinearity, which can subsequently have an effect on the accuracy of the model. Hence, PCA or ‘Principal Component Analysis’ was used to overcome this problem initially. PCA allows to convert the original variables into orthogonal variables (Brauner and Shacham, 2000). Hence, the original 26 variables were converted into 26 ‘Principal Components’ which were orthogonal variables. For the purpose of this study, it is beyond the scope to discuss how a Principal Component Analysis works. In simple terms however, the original variables were converted into new variables which are completely independent of each other. The eigen values signify the relationship between the principal components and the original variables. The eigen values were obtained which can be seen in table in appendix C.

As a rule of thumb, it is recommended to use the principal components with an eigen value of more than one (Wold, Esbensen and Geladi, 1987). Hence, from the output, it can be observed that the first nine principal components have an eigen value of more than one. In addition, these nine principal components can explain almost 80% of the variance in the original dataset. These principal components are linearly related to the original variables. The eigen vectors, which denote the relationship between the principal components and the original variables can be found in the appendix C. A regression analysis was then performed in MiniTab by replacing the original variables in the data matrix with the nine principal components. The results can be found in appendix C. A VIF value of 1 was obtained which signifies no collinearity issues, as was expected. However, a much lower R square value was obtained which signified that this model might be less effective at accounting for the variances that the previous approaches. In addition, this was also relatively more complicated to use as whenever the planner wishes to the use model for man-hour estimations, he would need to convert the original parameters of the operations (Uppercase alphabets in equation 2) to the principal components and thus, adding another step to the already complicated process of production planning. Time and again, it was pointed out that ease of use should be one of the important factors to be considered while carrying out this study to ensure successful adoption.
4.10 Approach 4: Average Values and Reduced Variance

One of the approaches for the ‘data cleansing’ process was to reduce the redundancy. Pan, Li and Yuan (2014) in their study had employed linear regression and had discussed on reformulating the input data for regression analysis. One of the approaches was considering average values. Upon a closer examination of the dataset, it was noted that in many entries, the ‘parameters’ of the operations were same, meaning, the exact same operations were performed for the same number of times for different subprojects. After a discussion with the production planner, it was observed that all of the windings were in-fact the exact same windings, just belonging to different subprojects. Hence, in order to reduce the variance in the dataset, the windings with the same parameters were combined into one entry. This was done by having the same independent variables on the right-hand side (RHS) of the equation (equation 2), while the average of the production times was taken with the end result being used for the left-hand side (LHS) of the equation (equation 2). This reduced the 250+ entries in the database to around 80+ entries and hence, reduced the variance in the dataset. Therefore, the total volume of data was also reduced. A high R square value was achieved, around 97.73%. This might be because of the less variance in the data and hence, it was ‘easier’ to find a ‘good fit’ for the data. Unfortunately, the problems with collinearity still persisted in addition to negative coefficients. The results of this approach from MiniTab can be found in Appendix D.

4.11 Approach 5: Increased Sample Size and Dimensionality Reduction

Multicollinearity being persistent in all the above approaches created an environment of actual presence of collinearity between the variables. Hence, correlation coefficients were checked between all the independent variables. Correlation coefficient is a standard measure of measuring the relationship between variables (Field, 2009). The results read that 10 variables were correlated with correlation coefficients reaching as high as 0.87, which implies, a variable is 87% positively correlated with another variable. As mentioned in approach 3 (Section 4.9), unstructured interviews had already been taken with production planner and production manager. But, they denied of any relationship between any variables (operations) and those operations were completely independent of each other. Field (2009) had suggested, collecting more data could help in reducing the multicollinearity. The sample size considered for previous approaches was more than 250. The sample size was calculated based on a thumb rule which stated for each predictor variable, there must be 10 to 15 datapoints (Field, 2009). This study consisted of 26 variables, and hence the estimated sample size was 260. This thumb rule was
criticised for oversimplifying the issue (ibid). The simplest thumb rule by Field (2009) was, bigger the sample size, the better. Moreover, to complement this thumb rule, Denis (2019) in his book described, increasing the sample size is usually the preferred method for boosting the statistical power of the model. Hence, to produce the best results, a new database was constructed with projects from 2015 to 2019, thus increasing the number of data records from 250 to more than 700. Moreover, by adding more data, more variances were included in the model, and in such industry where every project varies in engineering specifications, adding more data was justified. Once new data was added, data cleansing was done in a structured method as discussed earlier. Ordinary least square method (Section 4.7) was repeated with the new and extended database, and the results were relatively better to previous approaches and multicollinearity was reduced. Unfortunately, despite the multicollinearity issues being reduced, the issue still persisted. One of the initial thoughts was to add more data as it helped. However, the access to structured data was limited to the four years of production data, therefore, the possibility of adding more data could not be done. Field (2009) discusses another methodology to deal with multicollinearity and also mentions that there is no proper method to deal with multicollinearity – “unfortunately statistics is frustrating sometimes” (Field, 2009). Dimensionality reduction or reducing the number of predictor variables is another suggested methodology by Field (2009) for multicollinearity. Corrales, Corrales and Ledezma (2018) had also discussed about the importance of dimensionality reduction. Dimensionality reduction helps in reducing the multicollinearity issues (Schroeder, 1990). But Field (2009) discusses, there is no proper methodology to identify the variables that can be removed. In addition, since the results from the previous approaches were inconclusive, there was no way to know which ‘variables’ could be removed. This was an iterative process. Hence, dimensionality reduction, which is an approach of data cleansing, was again performed to remove some predictor variables. After many iterations, 6 variables were removed out of 26 predictor variables from analysis. This was done by bringing in the opinion of subject matter experts (SME). It should be noted that these operations were not ‘removed’ from the model, only from the analysis. Operation A, B and Z were the operations that were performed no matter the type of windings and hence, after a discussion with the production planner and the production operators, the coefficient value for these operations were fixed manually. The production planner, along with the production operators, estimated that operation A usually takes around 4 hours to complete, operation B takes around half an hour and operation Z takes around 4 hours again. In addition to these three variables, there were some other variables in the model that were repeatedly posing problems with either multicollinearity or negative coefficients. Even after adding all the
data that was available, the issues still persisted for some predictor variables. Hence, the variables D, F and L were also fixed manually. The production planner was interviewed to get his estimate on how much time it might take for these operations to be performed once. It should be noted however, that the initial plan was to calculate the time manually for problematic coefficients using an observational study or a time study and use this process in tandem with the process of incorporating the opinions of SME (Subject Matter Experts). However, the Covid-19 pandemic took away the possibility of travelling to the manufacturing plant. Hence, the only possibility that was left was the qualitative approach by incorporating the experience of experts. Once the coefficient values were fixed manually, it was important to modify the production times accordingly in the database as the production times still included the times for these operations. Since these operations were not going to be treated as predictor variables, running a regression analysis after removing them whilst the production time still accounts for them would have produced inaccurate results. For example, if operation A is performed once for a particular type of winding, since a coefficient value of 4 was fixed for this variable, 4 hours were subtracted from the total production time for every entry in the database that included operation A once (4*1=4). The same was done for all the operations and all the entries in the dataset before running a regression analysis in MiniTab with a lesser number of predictor variables. Upon doing so, the results that were achieved were (visually) far better than what had been achieved before. Along with non-zero positive coefficients, there was less multicollinearity in the model. Again, it is important to note that this was an iterative process involving different attempts to get a ‘visually’ suitable result (Less collinearity and no negative coefficients). The performance of the final result will be discussed further in the validation section 4.12. The R-square for the resulted model was 94.44%. The results of this approach can be found in Appendix E. The variables without VIF in the table are the variables that were removed from regression analysis. There are two variables in the table with relatively high VIF. This might be because in the dataset, the number of windings using these operations were relatively low and hence, statistically, there was not enough diversity in the data for these operations. They were still included however, as they contributed significantly to the total production time in the cases where they were used.

4.12 Validation

A mixture of qualitative and quantitative approach was performed for the validation process. In addition, owing to the time constraint of this study, it was not possible to validate the results of the regression analysis with future projects as the lead times are extremely long while the total
volume of units being manufactured at the winding department is extremely low. However, it was still important to carry out some validation as without doing so, it was not possible to answer the first research question regarding the usage of regression for production planning.

At first, a quantitative ‘data only’ approach was carried out for the model. After a suggestion received from the production planner and the production manager, some of the recently completed projects that included a wide variety of windings were studied including their total production times. The engineering parameters of the different operations for these windings were used as an input to the regression model and the final result was compared to the actual production times. A summary can be found in Table 8.

<table>
<thead>
<tr>
<th>Project</th>
<th>Sub Project</th>
<th>Winding Type</th>
<th>Actual Production Time</th>
<th>Predicted Time using Regression Model</th>
<th>Prediction Error</th>
<th>Prediction Error %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>Helical</td>
<td>61</td>
<td>88</td>
<td>27</td>
<td>44.26</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Disc</td>
<td>315</td>
<td>391</td>
<td>76</td>
<td>24.12</td>
</tr>
<tr>
<td></td>
<td>C1</td>
<td>Disc</td>
<td>234</td>
<td>300</td>
<td>66</td>
<td>28.20</td>
</tr>
<tr>
<td></td>
<td>C2</td>
<td>Disc</td>
<td>303</td>
<td>332</td>
<td>29</td>
<td>9.57</td>
</tr>
<tr>
<td>2</td>
<td>A</td>
<td>Helical</td>
<td>101</td>
<td>78</td>
<td>23</td>
<td>22.77</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Disc</td>
<td>314</td>
<td>284</td>
<td>30</td>
<td>9.55</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Helical</td>
<td>193</td>
<td>143</td>
<td>50</td>
<td>25.90</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>Disc</td>
<td>110</td>
<td>110</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>A</td>
<td>Disc</td>
<td>100</td>
<td>92</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>A</td>
<td>Disc</td>
<td>137</td>
<td>116</td>
<td>21</td>
<td>15.32</td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>Helical</td>
<td>69</td>
<td>61</td>
<td>8</td>
<td>11.59</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Disc</td>
<td>132</td>
<td>118</td>
<td>14</td>
<td>10.60</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Disc</td>
<td>96</td>
<td>132</td>
<td>36</td>
<td>37.5</td>
</tr>
<tr>
<td>7</td>
<td>A</td>
<td>Helical</td>
<td>57</td>
<td>64</td>
<td>7</td>
<td>12.28</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Disc</td>
<td>154</td>
<td>126</td>
<td>28</td>
<td>18.18</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Disc</td>
<td>114</td>
<td>133</td>
<td>19</td>
<td>16.66</td>
</tr>
</tbody>
</table>

Table 8: Validation of the resulted regression model against recently completed projects in the case company

The difference between the actual times and predicted times is around 18% on average. The implications of this will be discussed further in the analysis section.

At this stage, a qualitative approach for validation was also opted. Since the constant term was removed from the model, the total production time equalled to the product of number of times the respective operation was carried out multiplied with the time taken to complete that operation each time (coefficients obtained from the regression model). Hence, it enables the
possibility of discussing these coefficients with the production planners as well as the operators. Upon the discussion of the output with them, good feedback was received from the planner claiming that despite the lack of actual data to compare the coefficients, it seemed that the coefficients were ‘pretty close’ to the actual time it might take for those operations to be performed once, according to his experience. In addition, his discussion with the production operators yielded the same results.

Since there is a reliance on human intuition again, one might argue that there was no progress made as the end results were still relied on the experience of the experts. However, it should be noted that the validation process, which was carried out, used the experience of the experts in order to ‘aid’ the data, not to replace it completely. In addition, including the opinion of experts even aided in achieving the aforementioned results in the first place. Further exploration of this approach will be discussed in section 5.1 of this thesis.

4.13 Data-Driven Decision-Making in ETO

A major purpose of this study was to investigate the possibility of using data-driven tools such as regression for production planning purposes in an ETO company. As this study has repeatedly pointed out, there is a severe lack of production planning methodologies for ETO environments. In addition to methodologies, the ETO environment has also been largely ‘ignored’ by the academia. Hence, it is evident that no matter the scope of the study, the area still remains unexplored. Therefore, in order to investigate how ‘effective’ the production planning tool in focus for this study is, it might not be wise to benchmark it against those used in mass-manufacturing industry as the context is not the same. Both the industries operate under extremely different environments with different conditions. Therefore, in order to understand the effectiveness of this tool, a mixture of qualitative and quantitative approach was used. An unstructured interview was carried out with the production planner to understand the intricacies of production planning in the ETO environment and how this tool could prove useful. In addition, the validation process was carried out (explained previously) with the help of quantitative means as well as several unstructured interviews with the production planner to examine the ‘accuracy’ of the model.

The second research question of this study, however, focused on the successful adoption of such methods in the ETO industry. The production planner who had been actively involved in this study from the very beginning, was interviewed. He was chosen owing to his experience in the field of production planning in ETO industries. In addition, he had also been actively
involved in this study and was quite familiar with the regression model. The interview was semi-structured in nature. The interview guide can be found in the appendix F. In agreement with the confidentiality contract, it is not possible to disclose the transcription of the interview. Some key insights were obtained during this interview which will be further discussed in section 5.2.
5 ANALYSIS

This section answers the Research Questions through critical analysis of empirical findings

5.1 Research Question 1

How can regression analysis be used to estimate man-hours in an Engineering-To-Order environment?

This case study began with an in-depth review of the current literature on ETO environments, why production planning is important and why production planning can be extremely hard in ETO environments. While the empirical data did uncover some very interesting phenomena and possibilities of using regression for production planning in an ETO environment, the obstacles that were faced during this study were in-line with what was discussed in the literature. Before the question regarding ‘how’ can regression be used, it is first important to look at ‘if’ it can be used.

The ETO industry is associated with high levels of complexity and uncertainty in product specifications (Hicks and Braiden, 2000). In addition, there is also a high reliance on manual labour in ETO industries (Kumar and Shinde, 2019). It was interesting to see all of the aforementioned phenomena play out in the case company. The high levels of complexity in the product design combined with the uncertainty in the supply chain as well as the reliance on manual labour lead to a substantial amount of variance in the data itself. Unfortunately, it was not possible to examine closely on how this data was collected. However, during an unstructured interview with the production manager, the data that was used for this study was ‘reliable’ as a lot of focus had been put in to ensure that the data was collected properly. There were still some issues however, which will be discussed further. But the key point here is that if the data is to be trusted, there is a significant amount of variance in the data which might be because of the complications in the industry. Upon discussion with the production planner regarding such high variances during a semi-structured interview, the planner revealed that it was in-fact quite ‘realistic’ and ‘was to be expected’ during the production process. One of the major reasons that were discussed, was the difference in expertise levels of the production operators. According to the interviewee, some people have been working with the product for over ten years and having such experience can lead to faster production times if the ‘experienced’ operators are working during the shift. However, owing to the time constraint, it is not possible to account for all the variances. In addition, the production planner also pointed
out that changes can be made to a product being assembled mid-way through the process depending on the customer’s requirement. Hence, it is possible that these last-minute changes might have also contributed to the variance in the raw data. Interestingly, this was found in line with the literature on uncertainties in project planning in ETO industries. In fact, Vaagen, Kaut and Wallace (2017) discuss that one of the major obstacles that causes a lot complexity in project planning in the ETO industry is design changes that are made to the product during the engineering or the production process. This was pointed out as one of the major reasons for variances in the data by the production planner.

In addition to the aforementioned issue of variance, unavailability of data was another issue. Due to the Covid-19 pandemic, it was not possible to examine how the data collection process took place at the company. Hence, not much can be said about the data collection methodology and what effect it might have had on the study. This can be considered as one of the drawbacks of this study.

Nevertheless, the ability to examine this case in-depth still offered immense empirical data that enabled the possibility of investigating the use of regression for production planning in an ETO-based company. As was discussed before, it is first important to answer ‘if’ regression can be used. And the answer to this question is not as ‘black or white’. Whether regression can be used can depend on numerous factors. The accuracy levels that one might be looking for can affect whether regression can be used for production planning or not. As was seen in the validation section (Section 4.12), the mean absolute percentage difference between the actual and predicted time was almost 18%. One can argue that this can either be good or bad, depending on the context. When it was compared with the mass-manufacturing industry, the number seemed to be quite problematic. For example, a study that focuses on estimating costs for manufacturing a product in the mass-manufacturing industry created a model which resulted in a mean absolute percentage error of just 6.34% (Ning et al., 2020). However, it is important to consider the inherent differences between the two industries as was already highlighted in the literature review. A more interesting comparison is that of the software industry since the software industry also has ‘customized’ products (software in this case) depending on the preference of the customer. The software industry faces problems with cost-estimations as well. In fact, a study as far back as 1987 has noted concern over the inability to estimate costs in the software industry (Kemerer, 1987). It was quite a while ago and surely things must be different now? A number of different methodologies have evolved that focus on cost and ‘effort’ estimations in the software industry. A modelling process for cost and ‘effort’ estimation at
“Siemens Information Systems, India” yielded an error rate of about 65% in the first iteration, which subsequently reduced to about 55% in the second iteration (Trendowicz, 2013). Of course, this is not an ‘apples-to-apples’ comparison since the context is again, quite different. But it is interesting to see where that 18% lies on a spectrum. As discussed numerous times in this study, the ETO industry is still largely unexplored by literature when it comes to production planning (Carvalho, Oliveira and Scavarda, 2015). Putting things in context makes this revelation quite interesting and fascinating. Hence, for an industry that does not have a lot of production planning tools to rely on, this can be considered as an interesting first step towards a future with more data-driven decision making. However, there are still some very interesting findings to uncover regarding the ‘data-driven’ part of the decision-making process in this study that will be discussed subsequently.

The answer to the question regarding ‘how’ regression can be used involves different steps. Some of the most important findings of this study could be classified into three areas; data, different approaches to regression and SME (Subject Matter Experts).

The first area, data, has been in focus numerous times during this analysis with issues ranging from variance, unavailability as well as certainty regarding the accuracy of the data. As pointed out before, the ‘accuracy’ of the data was hard to study since the Covid-19 pandemic rendered it impossible to visit the factory and investigate how the data collection took place. Several unstructured interviews were taken in order to understand the process however, owing to the inability to travel, the possibility of examining the data collection methodology and possibly improving it was removed from the scope of the study.

One of the aspects of data that led to doubts regarding its inaccuracy was the variance in the data. As pointed out in the Empirics section, the variance in the total production time for the exact same winding could be as high as 50%. Even though the literature had repeatedly discussed the complications in the ETO industry, variances being one of them, it was still quite interesting to witness it first-hand. At this stage, a data cleansing methodology highlighted by Corrales, Corrales and Ledezma (2018) was followed to ensure to the best of abilities that the data was of ‘good’ quality. The procedure has already been highlighted in section 4.5. However, one of the interesting aspects is to consider the high levels of variance in the data. The data cleansing methodology by Corrales, Corrales and Ledezma (2018) highlights the importance of removing ‘outliers’. There are several algorithms that can assist in doing so. However, it should be noted that such high variances in the data can pose problems with this step. In
addition, it is important to consider that these variances are a part of the ETO industry as has been discussed by the literature (Section 2.2) and was also confirmed by the case company during the semi-structured interview with the production planner. So, if these variances are a part of reality, should they not be considered in the analysis? Unfortunately, the answer to this question is not a ‘Yes’ or a ‘No’. For the case company, the variances were extremely high in some cases and including such high variances reduced the statistical accuracy of the tests. After an unstructured interview with the production manager, it was noted that the management acknowledges the high variances and are not looking for perfect results. What this is, is merely a beginning towards a more data-driven approach to decision-making. However, as Adrodegari et al. (2015) pointed out, it is not possible to consider all ETO companies under the same umbrella. Hence, depending on the requirement, the answer to the question regarding including such variances in the data or not can differ.

The second area of analysis for this research question was mentioned as different approaches to regression. As was discussed in section 4.7 to section 4.11, several approaches were tried which eventually lead to several problems that were also described. However, this research question focuses on how regression can be used for man-hour estimations. Hence, it is important to ensure that the reasons for failure have been described in detail which will eventually assist in avoiding the same scenarios in the future. The very first problem that was encountered was that of negative coefficients. Yes, negative coefficients should not be practically possible since every operation that is performed just adds more time to the total production time. However, an interesting aspect to note was that the R square was still relatively high. Hence, it could be argued that the first approach could have predicted the total production time anyway. However, having negative coefficients signifies that there is a negative correlation between the respective independent variable and the dependent variable (Field, 2009), total production time in this case. That clearly cannot be in a practical situation. This opinion was held by the production planner as well. Hence, this approach was rejected. The second approach, which focused on forcing positive coefficients, yielded some coefficients as zero, which can also be never the case in a practical situation. After several discussions, it was noted that there might be a problem with the data that is hard to explain as in the first two approaches, a high VIF value was noted for several variables. Therefore, it was assumed that dealing with the high VIF values might yield better results, which is why PCA was adopted. However, as discussed, PCA was rejected as well. The first and foremost reason was that it offered the lowest R square value of all the approaches that had been tried. In addition, the
‘principal components’ are linearly related to the original variables. The original variables were the parameters that the planner interprets from the drawings. Hence, every time that the planner wishes to use the tool, he would need to convert the original parameters into the principal components and then run the analysis. This made the process extremely long and complicated and as will be explained subsequently for the second research questions, posed a problem in the adoption of the tool. Hence, this approach was rejected too. The fourth approach involved efforts in reducing the variances in the data. Similar problems were encountered which was why the last approach was opted for.

After all the approaches failed for different reasons, it was unclear as to what the root cause was. Clearly, relying on data wasn’t proving to be helpful. After a discussion with a production engineer, it was noted that she had encountered the same problems when she had tried to use regression as a production planning tool almost two years ago. During a brainstorming session, it was noted that bringing in the opinions of the production operators, the subject matter experts, might prove useful as they have been working in the industry for quite some time and knew the production processes in-depth. The production planner agreed with this idea. This was a turning point in the study. While this study started out as investigating more data-driven production planning tools in the ETO industry, it turned out that years of experience in an industry which deals with such high uncertainties and complexities cannot be negated after all. After several iterations of treating some variables as fixed, the coefficients of which were decided upon after several discussions with the production planner and the production operators, some of the best ‘visual’ results were achieved as there were no negative coefficients and the VIF values were relatively lower. The R square value was quite high as well. The process of validation merely confirmed that the final model was in-fact quite accurate considering the high variances in the data. Another study in the ship-building industry (an ETO industry) also achieved similar results (Wan Abd Rahman, Mhd Zaki and Abu Husain, 2019). Instead of completely replacing human-instincts with data, a more structured approach towards using data and ‘aiding’ it with the opinions of experts yielded great results (ibid). Unfortunately, due to the time constraint on this study, it is not possible to explore further as to the different possibilities of using data with human instincts.

In conclusion, the approach 5, which involved using both, data as well as human instincts, to ‘aid’ the decision-making process might be more effective. Hence, while working on production planning, it might be beneficial to look at both the sources of inputs as complements rather than substitutes.
5.2 Research Question 2

How can data-driven production planning tools such as regression analysis be successfully adopted in ETO industries?

The primary purpose of a data-driven production planning tool is to reduce the errors while planning and abiding by the planned schedule during the production. From the previous analysis, it can be said that regression analysis could be used for driving data-based decision in production planning, but that doesn’t necessarily implicate that this model would be adopted in the industry. Hence, adoption of a model in the industry would be the highest validation of that model. As discussed, ETO industries have huge variances in product specifications and labour knowledge and work style. Acknowledging these variances is very essential for the production planning tool to be adopted. Hence, during the semi-structured interview with the production planner, it was noted that the resulting production planning tool must be simple to understand and use, and accurately estimate the production hours.

Production planning based on regression analysis is simple to use and understand. In this tool the only input data is the number of times an operation is performed for a project, and with that single input, the tool could estimate the total production hours for the project. The production planner at the case company during an interview revealed that he would working closely with this model in the future for production planning. In addition, he also revealed that he found the model to be easy to use.

The production planning tool, while being simple to use, must also perform it’s intended function i.e. to produce accurate estimates. Accurate estimates could be provided if the production planning tool accounts various types of variances into it. From the previous analysis, it is evident that more than 90% of the variances in the case company had been accounted in the regression model. Hence, regression analysis produces estimates that should work with most of the projects as the regression solution is an average of datapoints. However, as discussed, acknowledging variances is very essential and in such industries, obtaining estimates that have high errors when compared with actual production time at shop floor is common and this was complemented by production planner as well as production executives in interviews. So, in order to improve the accuracy of the model, approach 5 which was performed for this study could be practised. Approach 5 (Section 4.11) discussed bringing in the opinions of subject matter experts into building the model. The subject matter experts for this study were the people who work in the production line. Bringing their opinion into the regression model, helped in
improving the accuracy of the model which is evident from the validation section of this study. The production planner in his interview also confirmed that the model looks “accurate”. This study could be misjudged for the model being biased towards shop floor, but it should be noted that, after interviews with shop floor operators, production leader and planner were also interviewed to confirm the values from the shop floor. Moreover, shop floor operators were chosen as subject matter experts in this study, because, the production line was relying upon manual labour for quite a long time and the operators have very deep knowledge on the production. In fact, during an unstructured interview with the production planner, it was revealed that it could take a minimum of five years for an operator to be considered ‘experienced’. Data driven decision making cannot be implemented suddenly in a production line that has been relying on operator experience for a long time. Hence, it was very essential to consider production operators’ opinion for this study which resulted in a positive impact. This study had many surprising turns, it started with exploring the techniques that were employed in the mass manufacturing industries for production planning, and regression was chosen to study as it was the primary model for many recent technologies. However, during this study, it was found that the regression that was used in mass manufacturing industries couldn’t be directly implemented for ETO industries. The regression needs to customized according to every company’s features. Hence, ETO industries cannot be compared with mass manufacturing or Make-to-stock sectors regarding the accuracy of the model. Mass manufacturing industries highly rely on machines for production with automated data collection. Hence more accurate data driven decision making is possible, which is not the case with ETO industries.

However, the accuracy of the model also depends on the input data. Though data cleansing was employed, the input data could still be improved. In addition, in the ETO industry, the variations in the specifications of the product are dependent on the customer, not the company. However, variations in the operation times might still be reduced through standardization, as was discussed by the production planner during an interview. This revelation was quite interesting as its implications had not yet been thought of during the course of the study. Standardization is a process of setting standards to prevent defects and occurrences of errors that could impact the production time (Mňkva et al., 2016). Despite being an essential key in the lean manufacturing (Mľkva et al., 2016) implementing standardization isn’t a simple task. In a study on implementation of standardized work in automotive industries in Sweden shows that truck manufacturers who customize their products for their customers are facing more difficulties
compared to car manufacturers (Johansson et al., 2013). Hence, a similar situation can be faced by the ETO industries. As discussed in the literature review, standardization could be a key for ETO industries in addressing the variances, while also maintaining the profitability and this was agreed by the production planner of the case company in an interview. However, the planner also commented that, standardization is possible in the case company.

The production planner mentioned that in a manual production line, standardization could be initiated by training people in a structured way. Currently, inexperienced operators get trained by spending time with an experienced operator, but every experienced operator has their own working way which again brings back the variance. Hence, as suggested by the interviewee, the training must be structured, in such a way that, every operator should be able to perform an operation at a similar pace which would result in same operation time. In a study conducted by Cannas et al. (2018) at a manual assembly line, suggested strategies such as Job Element Sheet (JES) and Standard Operating Sheet (SOS) for standardization. These are the documents that provide details on how to conduct a task to obtain desired time and quality (Cannas et al., 2018). Also, these documents can enable knowledge sharing between operators which is the first step towards performance improvement (Cannas et al., 2018) and it might even improve the learning curve of inexperienced operators as they might be spending relatively lesser time in learning while also learning the most efficient method. Another major factor derived from the interview that could implement standardization was the lack of involvement of upper management and the shop floor being hesitant to change. As discussed in section 2.10, employee acceptance plays a major role while implementing a significant organizational change. And, standardization is a major organizational change as a transition in employee’s work style is expected. In a study conducted at freight distribution company on the implementation process of standardisation by Gudmundsson, Boer and Corso (2004) revealed that lack of management support and resistance to change were some of the factors that posed hindrance to standardising processes.

One of the interesting perspectives to look at these issues simultaneously, could be labour union-management relation. Labour union is prevalent in the case company which can be the case with most ETO industries as the industry demands skilled manual labour due to complex product specifications. In a study conducted by Toribio and Hernández (2011) from Linnaeus University on coping with resistance to change in organizations, discussed, factors that could lead to resistance to change, some of the factors were group resistance and culture. Groups establish their own behavioural norms and they are prone to resisting the change if their
behavioural norms are getting disturbed (Toribio and Hernández, 2011). According to Cooper (1994) as cited in Toribio and Hernández (2011, p. 32), “culture based resistance is more prevalent when the direction of change and the handling methodology are not aligned with the cultural understandings and meaning characterizing the organization”. Also there had been numerous studies about union and management relations, a study by Dubin (1957) discusses about power and union-management relations and the significance of institutionalization of relations. The author concludes that union and management conflicts will exist even after institutionalizing, “every union-management relationship is always characterized by the simultaneous presence of elements of conflict and of cooperation” (Dubin, 1957). In a study conducted at a Japanese company which has labour union culture, revealed that, not only wages have been increased, but also the profitability and the productivity has increased, as the union worked in close relation with the management (Morikawa, 2010). Hence to conclude this discussion, ETO industries management could work in close cooperation with unions and employ their expertise to aid developments within the company.

The ultimate factor for successful adoption of a working model would be the acceptance of the model at all levels in the company. This factor could be the result of a cooperative union-management relationship. This study discusses the production planning, and hence for the empirical data, people from the production department - production manager and production engineer, people from planning function – production planner and people from the production line – shop floor operators were interviewed. As discussed, purposive sampling was done for this study and interviewees were chosen in such a way that the model gets accepted at all levels. A shop floor operator with around 30 years of experience in the case company was interviewed for the empirical data. The shop floor operators and winding shop production leaders seemed to be more enthusiastic while having unstructured interviews with them. Though this study is not about standardization of processes, it opens up the door for future standardizations that could be implemented in the case company which could be done by the collaborative effort between the management and the union.

Hence, to conclude, for a successful adoption of a regression-based production planning tool, it must be simple to use considering various other complexities in the industry, must be accurate in predictions, which is possible by including more data that is variant in nature as well as by reducing the differences in labour’s working style through standardization. And, standardization can be made possible by cooperative effort of union and management which also opens up door for higher acceptance level of the model.
6 DISCUSSION

This section discusses few topics that were of interest to this study.

6.1 Data-Collection

This study had discussed about the issues with data accuracy due to manual data collection and data reporting. Kumar and Shinde (2019) have discussed several methods for improving data collection from being unstandardized and manual based to standardised. Currently in the case company, the operators are filing the comments in ‘free format’ as termed in Kumar and Shinde (2019). Hence, instead of using a more broader way of expressing unplanned incidents, it could be narrowed down to a checklist format as suggested by Kumar and Shinde (2019). The checklist can contain a list of delays that could or had occurred in the shop floor in the past along with an option where the operator can explain the issue briefly if needed. This method helps in prioritising the issues and helps in finding the root cause by thorough analysis. Every operator has their own way of reporting comments; however, this method standardizes them which could be easier to analyse.

6.2 Union-Management relationship and Leadership needed in ETO

This study focussed on production planning, but also shed some light on labour union relationship with management as it was an inherent culture in the case company which is also one of the factors that could lead to a successful adoption of the production planning model. In the late 1930’s and 1940’s, the unionism was making huge progress, and often the management was in bitter terms with the union (Dubin, 1957). But analysts started observing a ‘change’ in relationship after long years of opposition, and later, the relationship started to shift towards ‘stability’ (ibid). According to Dubin (1957), the relationship was becoming institutionalized. And the author also discussed that the stable and institutionalized relationship can be enhanced by cooperative decision making such as joint time studies, joint welfare and health programs etc. One of the reasons for union-management relationship getting disturbed from a stable cooperation could be the resistance to change. The same factor was also identified from an empirical interview. This study had suggested standardization as a step towards improving the data accuracy and, implementing standardization in the company could lead to resistance of change, as standardization demands the job to be completed within the provided time by the management without compromising on the quality. Hence the present work culture is getting disturbed. In such cases, where group resistance is present and the change has to be
implemented to cope up with the market, transformational leadership could be a potential solution. Scholars like Bennis and Nanus (1985); Tichy and Devanna (1985) as cited in Toribio and Hernández (2011, p. 65), had argued about transformational leadership during transition period. According to Bass (1990) as cited in Toribio and Hernández (2011, p. 67), transformational leadership inspires the people to look beyond the self-interest for the good of the company by generating the awareness as well as acceptance of the company’s mission. Transformational leadership stirs interests by broadening and elevating the people’s motivation (ibid). Charismatic, Intellectual, Inspirational and considering Individual’s capability were some of characteristics of transformational leadership as highlighted in (Toribio and Hernández, 2011). A study was conducted by Twigg, Fuller and Hester (2008) on the effectiveness of transformational leadership with labour union leader and labour union members. The study concluded that transformational leadership increased the union-supportive behaviour as a ‘covenental’ relationship was fostered through this leadership style (Twigg, Fuller and Hester, 2008). Though the study by Twigg, Fuller and Hester (2008) doesn’t stress on union-management relationship, however the significant conclusion from the study was, transformational leadership increases supportive behaviour and fosters an agreement-based relationship.

6.3 Union-Management Trust

Union’s trust on management was one of the factors needs to be discussed for this study, as this factor can have significant effect on whether the shop floor would use this tool effectively and contribute for further efficiency enhancements. This could be the case for other ETO companies too which followed union culture, As discussed in the analysis section, this model could be improved further with more accurate data which was possible only if the union cooperates with the management. A study conducted on trust between union and management by Bryson (2001) concluded that

- Unions have higher trust on management when unions had sufficient power to challenge employers, where the union represented the union members effectively and where the unions were supported by the management. In such circumstances, the trust was higher or at least similar to non-unionised organizations.

- While, the management can also foster higher trust by ensuring that unions have sufficient power to make a positive contribution at the workplace
From the above two discussions, this study recommends that, conflicts are an essential component in union-management culture, but this should not be a hindrance towards the development of the company in this continuously evolving market. Active participation of management with the union, transformational leadership during transition periods and encouraging unions to actively participate in decision making could foster trust as well as build an institutionalized relationship.

6.4 Should ETO industries aim for complete data-driven decision-making?

The initial scope of this study was to examine the possibility of using data-driven decision-making tools such as regression for production planning in the ETO industry. However, after numerous attempts at it, the approach that produced the best results involved human instincts.

This can lead one to wonder, is completely data-driven decision making even possible in the ETO industry? After all, this study showcased how using data as an ‘aid’ instead of a replacement might be a good approach towards having more structured decision-making approaches. In addition, the study by Wan Abd Rahman, Mhd Zaki and Abu Husain (2019) also involved human instincts and subsequently, achieved good results. Due to a time constraint, it was not possible to explore this possibility, unfortunately. But there is still a strong reason to believe that complete data-driven decision making might not be possible in the ETO industry owing to complex design & the reliance on manual labour and subsequently, the variations that come with it. In addition, it was pointed out by Adrodegari et al. (2015) that the ETO industry is in dire need of production planning tools specifically customized for the industry. Hence, this can be an area of research for the ETO environment where the ‘data only’ approach is compared with ‘data and human instincts’.

6.5 Process Innovation

There exists a general equilibrium model that talks about the relationship between innovation and standardization and how they both are engines of growth, but they also compete with each other at the same time (Acemoglu, Gancia and Zilibotti, 2012). Innovation is often followed by standardization which ultimately results in wider adoption of new technologies and subsequently, lower costs (ibid). Before discussing any further, it is important to point out that this study does not conclude that this equilibrium model can somehow be associated with the empirical data and its analysis that was a part of this study. What this merely is, is a
As pointed out in the analysis of empirical data to answer RQ 2, standardization can be an important aspect of adopting data-driven production planning tools such as regression in ETO industries. One of the ways in which standardization might help in using regression efficiently would be to reduce the variation in raw data. However, there might be more implications to this process than what meets the eye. As per Acemoglu, Gancia and Zilibotti (2012), the process of innovation takes place first in the case of any new technology. However, during the initial stages, it is usually quite expensive and requires skilled personnel. Then, to assist in wide adoption of this new technology, the process of standardization takes place (ibid). Even though the situation described in this study is not directly comparable to the aforementioned phenomenon, it is still quite interesting to try and affix the results from the study on the spectrum of innovation and standardization.

According to Schilling (2013), a lot of the new technologies in the market exhibit what is called the ‘S curve’ when it comes to their performance improvement. This ‘S curve’ is plotting the performance improvement of the said technology versus the time or the effort that is put in to make this technology better over a period (ibid). The reason why it’s called the ‘S curve’ is because it follows the shape of a horizontal ‘S’. The performance improvements usually start quite slow in the beginning requiring a lot of effort or money. Eventually, however, the rate of performance improvements increases (ibid). Exploring this phenomenon is not within the bounds of this study. However, comparing this to the study by Acemoglu, Gancia and Zilibotti (2012) as some very similar things were discussed.

Acemoglu, Gancia and Zilibotti (2012) also discussed how innovation can be expensive in the beginning and require skilled personnel but standardization eventually assists in bringing costs down. Eventually, it also results in lowering the bar of ‘entry’ in a way that workers of almost any skill level can perform the production operations. This is again similar to Schilling’s (2013) description of the ‘S Curve’ where eventually, the cost and the effort to get performance improvement go down. Another interesting aspect to note is that innovation can be of different types. Discussing all the different types of innovation is not relevant to this study. However, Schilling (2013) discusses how ‘process innovation’, a type of innovation, involves changing how an organization goes by conducting its business. Therefore, since this study involves
investigating new ways of production planning, it can directly affect the way the case company goes about with its production planning process.

Hence, if this new way of production planning is successfully implemented in the case company, it can be said that as of the current state, it is at the beginning of the S curve. A lot of effort needs to be put in in order to make improvements to the process. However, as was discussed in the analysis section for the second research question, standardization might further assist in making the regression model more accurate as well as help in successful adoption of the model. In addition, if this projection of the current situation on the ‘S’ curve described by Schilling (2013) turns out to be accurate, it can also be said that eventually, the effort it requires to make this process innovation better might go down. In addition, as Acemoglu, Gancia and Zilibotti (2012) described, increased standardization might also lead to reduced costs.

The implications of this projection are unknown. However, it can be still be very interesting to look at the current production planning scenario in the ETO industry and examine it from the perspective of ‘process innovation’ as it might assist ETO companies in drawing out a future ‘roadmap’. In addition, it might also help companies in knowing what to expect as it can disheartening to put in the effort and yet, yield minimal results.

6.6 Challenges during the study

Due to the global pandemic situation, access and travel were restricted. And, accordingly the research schedule had to be changed. As discussed, there were plans of conducting empirical observational study, but later, the plan had to be cancelled. Also, lot of in-person meetings and interviews were converted to online meetings and interviews, which hadn’t reduced efficiency, but, in-person interviews would have been much effective over online interviews.
7 CONCLUSION

This section concludes the thesis along with a brief discussion of how this study contributes to the research world and how it could be employed for further research

7.1 Concluding the study

Nowadays, the businesses are evolving rapidly, and the industries are gearing up for the fourth industrial revolution. The manufacturing industries are investing heavily in digitalization, machine learning, AI etc. The desire to move towards errorless data-based predictive analytics is continuously increasing. This was the driving force for this study. In a market, where mass manufacturers have successfully implemented advanced technologies to predict costs, customer behaviour, market fluctuations etc., Engineering-to-Order companies face difficulties in implementing the similar technologies due to various complexities involved in the industry. Due to which, these industries are still basing their decisions on instincts and expertise. The significant factor in cost estimation for ETO industries relying on manual labour is the labour cost of a project and predicting it could serve in bidding process as well as for resource planning. Hence, instead of directly investigating the most advanced technologies, this research studies the possibilities of data driven predictions through regression-based tool, which is the powerhouse of many advanced prediction technologies, in ETO industries for estimating man-hours of a project.

This thesis answers the possibilities of using a regression-based production planning tool in ETO industries through two research questions. RQ1 focusses on how the regression model can be made to operate effectively in an ETO environment. RQ1 concludes that, the regression-based tool can be employed in the ETO industries by aiding the results based on historic data with the opinions of SME. This study had also validated the resulted model against recently completed projects and found that the average error lies at nominal distance from the mass manufacturing industries in the spectrum considering variances and complexities naturally associated with the industry. While, RQ2 focusses on how the resulting model can be successfully adopted in the industry. The motive behind including this research question in the study was to ensure that the output was not just ‘statistically effective’ but also practically feasible and usable. Hence, to answer this research question, factors for successful adoption from empirical findings were analysed. RQ1 had shed some light on the huge variances in data that affected the accuracy of the regression model, while this issue has been discussed in detail
with RQ2. Strategies such as standardization and union-management cooperation are discussed which can lead to lesser variances in the dataset as well as higher acceptance of the model in the industry.

### 7.2 Academic and Practical Contributions

As discussed, there aren’t much research conducted on production planning in ETO industries specifically on man-hour estimations, this study will serve as a first step towards a more structured approach to deal with production planning in industries with many tangible and intangible complexities. A case study research design makes this study difficult to generalize and validate. But much emphasis has been put into replicating the study through a detailed explanation of various approaches to regression analyses.

Despite of difficulties in generalization, this study can still be employed in few situations with similar characteristics:

1. High variance in the production data
2. High reliance on the manual labour
3. Presence of Union-Management culture
4. Lack of standardization in operation

In the aforementioned situations, this study can be used to gain insights on employing regression for man-hour estimations, however, it should be noted that, the variances vary among the industries, hence, required customizations must be done. However, as discussed, this study paves way for industries aiming for the adoption of a structured approach towards production planning by basing the prediction on data rather than a traditional approach in ETO environment.

### 7.3 Further Research

As discussed, this research is just a ‘starting point’ for production planning methodologies in the ETO industry. Production Planning is not merely estimating man-hours for an operation or for a product, it involves numerous other factors as well. Hence, it might be interesting to look at what these other factors can be for an ETO industry to have an efficient and effective production planning process. In addition, owing to the time constraint, only regression was used as a tool to estimate the man-hours. Machine learning techniques, which are based on regression, might be able to cope with all the variances in the ETO industry and hence, the next
step in the process might be to look at how ML techniques can be used for man-hour estimations in the ETO industry. In addition, it should be noted that not all the variances could be accounted for in this model. Standardisation was proposed as a solution to some of the problems with the variances. However, implementing standardisation is not an easy job. Standardisation in the ETO industry could be another area of research as it opens up immense possibilities of reducing complexities which are normally associated with the ETO industry.

In addition, this study was performed at just one ETO company owing to the requirement and processing of a great amount of data. This could cause issues with validity as have been discussed. Performing similar studies at other ETO companies could greatly improve the validity of using regression in the ETO industry.
REFERENCES


Blomkvist, P. and Hallin, A., 2015. Method for engineering students: degree projects using the 4-phase model. 1. ed. Available at: <http://uu.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwdV09b8lwED0hWNigH3y1xVM36JHg6DJWCFQh0qITF-QmF4mFxQri53OXRARVYrNlybKtk9_d2e8dQBJMcfbvTogdZkipys9lhH8LzKXtCJkX4n9YJScnv_T5TduE1s1_2ezgnVbD0d8XJU2Nz_70oW6J-PmBqoCGErOLFc_tSkKuWF8qYysIWRnNH0lORY3wLHp>.


Carvalho, A.N., Oliveira, F. and Scavarda, L.F., 2015. Tactical capacity planning in a real-


Kumar, R. and Shinde, S., 2019. Simulation as decision support tool A case study with data analysis , VSM and simulation applied to an ETO System. Uppsala Universitet.


Lennerfors, T.T., 2019. Ethics in engineering. First ed. [online] Lund: Studentlitteratur. Available at: <http://uu.summon.serialssolutions.com/2.0.0/link/0/eLvHCXMwdZ1LC81wDMtD6MWbT3yL6CMdt3Wo8hEBMWDJy-Sbi488IT_fY2heC1x1AobSLN_iW_BeCKqT_5eRMy6VvMvJTrmTROG1aLWNJnJMG D8eb2PZ5t4tY6TT75sdwjd8Ph7IsHptbX84UJbsGL2N70ml2d4unau4kll2Z1fArYrSOm-b203m-ZfjWNSgzDBB>.


Mauergauz, Y., 2016. *Advanced planning and scheduling in manufacturing and supply chains.* *Advanced Planning and Scheduling in Manufacturing and Supply Chains.*


Toribio, C.T. and Hernández, R.G., 2011. Coping with Resistance to Change in Organizations. *Trendowicz, A., 2013. Software cost estimation, benchmarking, and risk assessment: the software decision-makers’ guide to predictable software development*. 1. Aufl.;2 ed. [online] Heidelberg;New York; Springer. Available at: <http://uu.summon.serialssolutions.com/2.0.0/link/0/elvHCXMwDV1LT8MwDLbGuMCFt1Y2WMR5nULSRyI0CKZNCDOqC0ISv0HiLGhtYW_j5OuXTtp6qGK7DZy2jiOY38G4KxP7Q2dgHaFEwkRengxvAlJIO-4bhDiqfSmOt95_CbunsXjWAWzWrdFMdmmvQz6dWx0pMivj44hDT6PZV3anb9zWCQblh4jYa3Nrh4nFb7OBejo4>.


Zhong, R.Y., Dai, Q.Y., Qu, T., Hu, G.J. and Huang, G.Q., 2013. RFID-enabled real-time

Appendix A: Results from MiniTab using Ordinary Least Square Method

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Coefficient</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-21.3</td>
<td>28.79</td>
</tr>
<tr>
<td>B</td>
<td>-17.77</td>
<td>8.13</td>
</tr>
<tr>
<td>C</td>
<td>0.872</td>
<td>23.75</td>
</tr>
<tr>
<td>D</td>
<td>-1.66</td>
<td>25.48</td>
</tr>
<tr>
<td>E</td>
<td>2.839</td>
<td>4.46</td>
</tr>
<tr>
<td>F</td>
<td>4.62</td>
<td>15.01</td>
</tr>
<tr>
<td>G</td>
<td>1.998</td>
<td>1.95</td>
</tr>
<tr>
<td>H</td>
<td>4.88</td>
<td>6.57</td>
</tr>
<tr>
<td>I</td>
<td>0.1041</td>
<td>37.86</td>
</tr>
<tr>
<td>J</td>
<td>-0.411</td>
<td>2.26</td>
</tr>
<tr>
<td>K</td>
<td>0.662</td>
<td>3.62</td>
</tr>
<tr>
<td>L</td>
<td>-3.895</td>
<td>3.00</td>
</tr>
<tr>
<td>M</td>
<td>0.043</td>
<td>4.27</td>
</tr>
<tr>
<td>N</td>
<td>0.03568</td>
<td>34.16</td>
</tr>
<tr>
<td>O</td>
<td>0.00781</td>
<td>4.80</td>
</tr>
<tr>
<td>P</td>
<td>0.0569</td>
<td>167.19</td>
</tr>
<tr>
<td>Q</td>
<td>-1.12</td>
<td>1.86</td>
</tr>
<tr>
<td>R</td>
<td>3.16</td>
<td>10.48</td>
</tr>
<tr>
<td>S</td>
<td>1.171</td>
<td>2.64</td>
</tr>
<tr>
<td>T</td>
<td>-0.740</td>
<td>31.73</td>
</tr>
<tr>
<td>U</td>
<td>-1.800</td>
<td>139.41</td>
</tr>
<tr>
<td>V</td>
<td>9.7</td>
<td>4.34</td>
</tr>
<tr>
<td>W</td>
<td>36.72</td>
<td>2.27</td>
</tr>
<tr>
<td>X</td>
<td>17.9</td>
<td>7.07</td>
</tr>
<tr>
<td>Y</td>
<td>0.391</td>
<td>5.09</td>
</tr>
<tr>
<td>Z</td>
<td>-28.25</td>
<td>3.77</td>
</tr>
</tbody>
</table>
## Appendix B: Results from R using NNLS method

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Coefficients</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.0000000000</td>
<td>10.733513</td>
</tr>
<tr>
<td>B</td>
<td>0.0000000000</td>
<td>3.800325</td>
</tr>
<tr>
<td>C</td>
<td>1.713940906</td>
<td>15.528138</td>
</tr>
<tr>
<td>D</td>
<td>0.0000000000</td>
<td>22.817835</td>
</tr>
<tr>
<td>E</td>
<td>1.739639976</td>
<td>4.707743</td>
</tr>
<tr>
<td>F</td>
<td>0.646897258</td>
<td>8.941333</td>
</tr>
<tr>
<td>G</td>
<td>2.070870874</td>
<td>2.237069</td>
</tr>
<tr>
<td>H</td>
<td>4.63920236</td>
<td>6.812276</td>
</tr>
<tr>
<td>I</td>
<td>0.066986258</td>
<td>21.675414</td>
</tr>
<tr>
<td>J</td>
<td>0.703301843</td>
<td>1.980699</td>
</tr>
<tr>
<td>K</td>
<td>2.421974401</td>
<td>5.359688</td>
</tr>
<tr>
<td>L</td>
<td>0.0000000000</td>
<td>3.341582</td>
</tr>
<tr>
<td>M</td>
<td>1.723657931</td>
<td>4.815555</td>
</tr>
<tr>
<td>N</td>
<td>0.021587172</td>
<td>23.375036</td>
</tr>
<tr>
<td>O</td>
<td>0.006211645</td>
<td>4.237922</td>
</tr>
<tr>
<td>P</td>
<td>0.010479774</td>
<td>160.480360</td>
</tr>
<tr>
<td>Q</td>
<td>1.459231424</td>
<td>1.812838</td>
</tr>
<tr>
<td>R</td>
<td>0.0000000000</td>
<td>10.856463</td>
</tr>
<tr>
<td>S</td>
<td>0.492998012</td>
<td>3.271433</td>
</tr>
<tr>
<td>T</td>
<td>0.128224456</td>
<td>31.030641</td>
</tr>
<tr>
<td>U</td>
<td>0.0000000000</td>
<td>137.529044</td>
</tr>
<tr>
<td>V</td>
<td>72.038422897</td>
<td>4.265518</td>
</tr>
<tr>
<td>W</td>
<td>38.939681265</td>
<td>2.580567</td>
</tr>
<tr>
<td>X</td>
<td>2.688037772</td>
<td>7.365359</td>
</tr>
<tr>
<td>Y</td>
<td>0.342249510</td>
<td>3.299523</td>
</tr>
<tr>
<td>Z</td>
<td>0.0000000000</td>
<td>1.982029</td>
</tr>
</tbody>
</table>
### Appendix C: Eigen analysis of the Correlation Matrix

<table>
<thead>
<tr>
<th>Eigen value</th>
<th>Proportion</th>
<th>Cumulative</th>
<th>Eigen value</th>
<th>Proportion</th>
<th>Cumulative</th>
<th>Eigen value</th>
<th>Proportion</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.424</td>
<td>0.209</td>
<td>0.209</td>
<td>3.491</td>
<td>0.134</td>
<td>0.343</td>
<td>2.245</td>
<td>0.086</td>
<td>0.535</td>
</tr>
<tr>
<td>2.739</td>
<td>0.134</td>
<td>0.343</td>
<td>1.756</td>
<td>0.068</td>
<td>0.448</td>
<td>1.607</td>
<td>0.062</td>
<td>0.565</td>
</tr>
<tr>
<td>2.145</td>
<td>0.105</td>
<td>0.448</td>
<td>1.386</td>
<td>0.046</td>
<td>0.502</td>
<td>1.19</td>
<td>0.046</td>
<td>0.635</td>
</tr>
<tr>
<td>1.142</td>
<td>0.068</td>
<td>0.514</td>
<td>1.19</td>
<td>0.044</td>
<td>0.506</td>
<td>1.142</td>
<td>0.044</td>
<td>0.653</td>
</tr>
<tr>
<td>0.988</td>
<td>0.038</td>
<td>0.552</td>
<td>1.000</td>
<td>0.005</td>
<td>0.999</td>
<td>0.988</td>
<td>0.005</td>
<td>0.999</td>
</tr>
</tbody>
</table>
Appendix C: Eigen Vectors of nine principal components

<table>
<thead>
<tr>
<th>Variable</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
<th>PC7</th>
<th>PC8</th>
<th>PC9</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-0.093</td>
<td>-0.171</td>
<td>-0.296</td>
<td>0.025</td>
<td>-0.295</td>
<td>-0.487</td>
<td>-0.001</td>
<td>-0.019</td>
<td>-0.018</td>
</tr>
<tr>
<td>B</td>
<td>-0.254</td>
<td>-0.197</td>
<td>-0.191</td>
<td>0.095</td>
<td>-0.233</td>
<td>-0.101</td>
<td>0.141</td>
<td>-0.2</td>
<td>-0.02</td>
</tr>
<tr>
<td>C</td>
<td>-0.344</td>
<td>0.08</td>
<td>-0.111</td>
<td>0.034</td>
<td>0.177</td>
<td>0.155</td>
<td>-0.239</td>
<td>0.184</td>
<td>0.041</td>
</tr>
<tr>
<td>D</td>
<td>-0.049</td>
<td>-0.446</td>
<td>-0.049</td>
<td>-0.22</td>
<td>0.019</td>
<td>-0.181</td>
<td>0.144</td>
<td>-0.141</td>
<td>-0.022</td>
</tr>
<tr>
<td>E</td>
<td>0.023</td>
<td>-0.41</td>
<td>0.028</td>
<td>-0.239</td>
<td>0.125</td>
<td>0.17</td>
<td>-0.058</td>
<td>0.071</td>
<td>0.009</td>
</tr>
<tr>
<td>F</td>
<td>-0.343</td>
<td>-0.056</td>
<td>-0.164</td>
<td>-0.034</td>
<td>0.092</td>
<td>-0.097</td>
<td>-0.126</td>
<td>0.011</td>
<td>0.057</td>
</tr>
<tr>
<td>G</td>
<td>0.093</td>
<td>-0.021</td>
<td>0.038</td>
<td>0.193</td>
<td>-0.02</td>
<td>0.036</td>
<td>0.202</td>
<td>0.504</td>
<td>0.255</td>
</tr>
<tr>
<td>H</td>
<td>0.276</td>
<td>0.086</td>
<td>-0.346</td>
<td>-0.026</td>
<td>0.09</td>
<td>0.064</td>
<td>-0.233</td>
<td>-0.14</td>
<td>-0.049</td>
</tr>
<tr>
<td>I</td>
<td>-0.402</td>
<td>0.05</td>
<td>-0.08</td>
<td>0.009</td>
<td>0.149</td>
<td>-0.007</td>
<td>-0.057</td>
<td>0.031</td>
<td>0.021</td>
</tr>
<tr>
<td>J</td>
<td>0.046</td>
<td>0.062</td>
<td>0.248</td>
<td>-0.174</td>
<td>-0.234</td>
<td>-0.259</td>
<td>-0.096</td>
<td>0.389</td>
<td>-0.21</td>
</tr>
<tr>
<td>K</td>
<td>0.034</td>
<td>0.098</td>
<td>0.324</td>
<td>-0.085</td>
<td>0.244</td>
<td>0.255</td>
<td>0.31</td>
<td>-0.293</td>
<td>-0.025</td>
</tr>
<tr>
<td>L</td>
<td>-0.162</td>
<td>0.054</td>
<td>-0.01</td>
<td>-0.013</td>
<td>0.13</td>
<td>-0.026</td>
<td>-0.185</td>
<td>0.008</td>
<td>0.754</td>
</tr>
<tr>
<td>M</td>
<td>-0.253</td>
<td>0.065</td>
<td>-0.08</td>
<td>0.019</td>
<td>0.123</td>
<td>0.181</td>
<td>-0.151</td>
<td>0.211</td>
<td>-0.525</td>
</tr>
<tr>
<td>N</td>
<td>-0.373</td>
<td>0.087</td>
<td>0.104</td>
<td>-0.047</td>
<td>0.097</td>
<td>-0.085</td>
<td>0.142</td>
<td>-0.028</td>
<td>-0.071</td>
</tr>
<tr>
<td>O</td>
<td>0.224</td>
<td>0.082</td>
<td>-0.37</td>
<td>-0.033</td>
<td>0.261</td>
<td>-0.013</td>
<td>-0.028</td>
<td>-0.049</td>
<td>-0.087</td>
</tr>
<tr>
<td>P</td>
<td>0.059</td>
<td>-0.474</td>
<td>0.098</td>
<td>0.089</td>
<td>0.152</td>
<td>0.119</td>
<td>-0.138</td>
<td>0.065</td>
<td>-0.013</td>
</tr>
<tr>
<td>Q</td>
<td>0.026</td>
<td>-0.107</td>
<td>0.086</td>
<td>0.4</td>
<td>-0.034</td>
<td>0.05</td>
<td>0.033</td>
<td>-0.025</td>
<td>-0.062</td>
</tr>
<tr>
<td>R</td>
<td>0.269</td>
<td>0.09</td>
<td>-0.373</td>
<td>-0.011</td>
<td>0.238</td>
<td>-0.001</td>
<td>0.038</td>
<td>0.15</td>
<td>0.015</td>
</tr>
<tr>
<td>S</td>
<td>0.045</td>
<td>-0.006</td>
<td>-0.063</td>
<td>-0.005</td>
<td>-0.509</td>
<td>0.359</td>
<td>-0.012</td>
<td>-0.277</td>
<td>0.093</td>
</tr>
<tr>
<td>T</td>
<td>0.071</td>
<td>-0.157</td>
<td>0.132</td>
<td>0.524</td>
<td>0.048</td>
<td>-0.039</td>
<td>-0.218</td>
<td>-0.049</td>
<td>-0.049</td>
</tr>
<tr>
<td>U</td>
<td>0.028</td>
<td>-0.45</td>
<td>0.035</td>
<td>-0.24</td>
<td>0.147</td>
<td>0.162</td>
<td>-0.068</td>
<td>0.076</td>
<td>0.008</td>
</tr>
<tr>
<td>V</td>
<td>0.036</td>
<td>-0.14</td>
<td>0.092</td>
<td>0.475</td>
<td>0.044</td>
<td>-0.029</td>
<td>-0.023</td>
<td>0.083</td>
<td>-0.003</td>
</tr>
<tr>
<td>W</td>
<td>0.059</td>
<td>0.01</td>
<td>0.09</td>
<td>0.017</td>
<td>0.008</td>
<td>-0.11</td>
<td>-0.612</td>
<td>-0.308</td>
<td>-0.047</td>
</tr>
<tr>
<td>X</td>
<td>0.042</td>
<td>0.133</td>
<td>0.428</td>
<td>-0.18</td>
<td>0.083</td>
<td>-0.237</td>
<td>-0.261</td>
<td>-0.106</td>
<td>0.009</td>
</tr>
<tr>
<td>Y</td>
<td>-0.233</td>
<td>0.032</td>
<td>-0.088</td>
<td>0.048</td>
<td>-0.226</td>
<td>0.444</td>
<td>-0.093</td>
<td>-0.032</td>
<td>-0.033</td>
</tr>
<tr>
<td>Z</td>
<td>-0.118</td>
<td>0.011</td>
<td>-0.016</td>
<td>0.205</td>
<td>0.352</td>
<td>-0.2</td>
<td>0.275</td>
<td>-0.338</td>
<td>-0.076</td>
</tr>
</tbody>
</table>
## Appendix C: PCR Results from MiniTab

<table>
<thead>
<tr>
<th>Term</th>
<th>Coefficient</th>
<th>SE Coefficient</th>
<th>T-Value</th>
<th>P-Value</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0</td>
<td>3.25</td>
<td>0</td>
<td>0.999</td>
<td></td>
</tr>
<tr>
<td>PC1</td>
<td>-34.48</td>
<td>1.4</td>
<td>-24.69</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PC2</td>
<td>-3.1</td>
<td>1.74</td>
<td>-1.78</td>
<td>0.077</td>
<td>1</td>
</tr>
<tr>
<td>PC3</td>
<td>-0.42</td>
<td>1.96</td>
<td>-0.21</td>
<td>0.831</td>
<td>1</td>
</tr>
<tr>
<td>PC4</td>
<td>4.81</td>
<td>2.17</td>
<td>2.21</td>
<td>0.028</td>
<td>1</td>
</tr>
<tr>
<td>PC5</td>
<td>23.1</td>
<td>2.46</td>
<td>9.41</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PC6</td>
<td>22.53</td>
<td>2.57</td>
<td>8.78</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PC7</td>
<td>-29.49</td>
<td>2.76</td>
<td>-10.67</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>PC8</td>
<td>2.43</td>
<td>2.98</td>
<td>0.81</td>
<td>0.416</td>
<td>1</td>
</tr>
<tr>
<td>PC9</td>
<td>-16.1</td>
<td>3.04</td>
<td>-5.29</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Appendix D: Results from MiniTab through average inputs

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Coefficients</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-2.3</td>
<td>31.61</td>
</tr>
<tr>
<td>B</td>
<td>-11.1</td>
<td>8.06</td>
</tr>
<tr>
<td>C</td>
<td>0.28</td>
<td>32.5</td>
</tr>
<tr>
<td>D</td>
<td>-1.36</td>
<td>40.69</td>
</tr>
<tr>
<td>E</td>
<td>2.765</td>
<td>7.06</td>
</tr>
<tr>
<td>F</td>
<td>2.76</td>
<td>15.48</td>
</tr>
<tr>
<td>G</td>
<td>1.57</td>
<td>1.94</td>
</tr>
<tr>
<td>H</td>
<td>4.89</td>
<td>7.55</td>
</tr>
<tr>
<td>I</td>
<td>0.1566</td>
<td>30.5</td>
</tr>
<tr>
<td>J</td>
<td>0.061</td>
<td>2.03</td>
</tr>
<tr>
<td>K</td>
<td>1.74</td>
<td>2.68</td>
</tr>
<tr>
<td>L</td>
<td>-2.47</td>
<td>6.92</td>
</tr>
<tr>
<td>M</td>
<td>0.698</td>
<td>6.17</td>
</tr>
<tr>
<td>N</td>
<td>0.0246</td>
<td>34.3</td>
</tr>
<tr>
<td>O</td>
<td>0.005</td>
<td>6.6</td>
</tr>
<tr>
<td>P</td>
<td>0.0559</td>
<td>315.04</td>
</tr>
<tr>
<td>Q</td>
<td>-1.14</td>
<td>3.2</td>
</tr>
<tr>
<td>R</td>
<td>1.8</td>
<td>10.46</td>
</tr>
<tr>
<td>S</td>
<td>0.835</td>
<td>1.57</td>
</tr>
<tr>
<td>T</td>
<td>-0.825</td>
<td>37.01</td>
</tr>
<tr>
<td>U</td>
<td>-1.85</td>
<td>306.91</td>
</tr>
<tr>
<td>V</td>
<td>5.3</td>
<td>2.36</td>
</tr>
<tr>
<td>W</td>
<td>38.03</td>
<td>2.56</td>
</tr>
<tr>
<td>X</td>
<td>8.2</td>
<td>6.41</td>
</tr>
<tr>
<td>Y</td>
<td>0.22</td>
<td>6.57</td>
</tr>
<tr>
<td>Z</td>
<td>-26</td>
<td>5.1</td>
</tr>
</tbody>
</table>
Appendix E: Results from MiniTab by combining Ordinary Least Square method and Qualitative feedback

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Coefficients</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1.47</td>
<td>5.14</td>
</tr>
<tr>
<td>D</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>1.22</td>
<td>39.57</td>
</tr>
<tr>
<td>F</td>
<td>2.83</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>2.73</td>
<td>1.31</td>
</tr>
<tr>
<td>H</td>
<td>4.09</td>
<td>5.41</td>
</tr>
<tr>
<td>I</td>
<td>0.07</td>
<td>8.70</td>
</tr>
<tr>
<td>J</td>
<td>0.941</td>
<td>1.21</td>
</tr>
<tr>
<td>K</td>
<td>2.8</td>
<td>1.39</td>
</tr>
<tr>
<td>L</td>
<td>1.18</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>1.46</td>
<td>1.64</td>
</tr>
<tr>
<td>N</td>
<td>0.017</td>
<td>8.41</td>
</tr>
<tr>
<td>O</td>
<td>0.005</td>
<td>2.00</td>
</tr>
<tr>
<td>P</td>
<td>0.0058</td>
<td>2.56</td>
</tr>
<tr>
<td>Q</td>
<td>2.55</td>
<td>2.36</td>
</tr>
<tr>
<td>R</td>
<td>1.52</td>
<td>3.72</td>
</tr>
<tr>
<td>S</td>
<td>0.48</td>
<td>1.03</td>
</tr>
<tr>
<td>T</td>
<td>0.23</td>
<td>3.95</td>
</tr>
<tr>
<td>U</td>
<td>0.2</td>
<td>39.02</td>
</tr>
<tr>
<td>V</td>
<td>39.9</td>
<td>1.41</td>
</tr>
<tr>
<td>W</td>
<td>38.68</td>
<td>1.66</td>
</tr>
<tr>
<td>X</td>
<td>15.76</td>
<td>1.14</td>
</tr>
<tr>
<td>Y</td>
<td>0.32</td>
<td>2.07</td>
</tr>
<tr>
<td>Z</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>
Appendix F: Interview Guide

[Introduction]

- What is your role and your responsibilities in production planning?
- Can you explain in brief the production environment in this company for instance, products manufactured, production processes, repeating issues etc.
- How long have you been working in a similar environment?

[Main Body]

- How does a process start once you have received a proposal? (Proposal to Order)
- How does project proceeds after initiating?
- At which point, does your role come in and how do you do it?
- In general, what is your view on the current production planning process at the company?
- How do you create the production plan?
- What are some of the common planning tools and how satisfied are you with those?
- How does the planning tool cope up with the variances in the design specs?
- How do you deal with uncertainties such as the differences in the manual labour etc.?
- How well is the production planning process and tools accepted at the company?
  a. How do you think its acceptance levels could be improved?
  b. How important do you think it is for the process and tools for production planning to be accepted?

[Exit]

- Is there anything we missed, and you would like to add up that could help with this study?