



UNIVERSIDADE CATÓLICA PORTUGUESA

Assigning Due Dates to Jobs in a Manufacturing Company

José Maria Antunes Bento Moreira

Católica Porto Business School

2023



UNIVERSIDADE CATÓLICA PORTUGUESA

Assigning Due Dates to Jobs in a Manufacturing Company

Master's Final Work in the modality of Dissertation presented to Universidade Católica Portuguesa to fulfill the requirements for the degree of Master in Management

by

José Maria Antunes Bento Moreira

under the orientation of
Aydin Teymourifar, PHD

Católica Porto Business School, Universidade Católica Portuguesa
April, 2023

Acknowledgments

In the first place, I would like to thank my supervisor, Aydin Teymourifar, for his guidance, encouragement, and dedication throughout this journey, which challenged me to think critically and to make me a better student.

To my family, especially my mother, grandmother, and brother, thank you. Thank you for taking care of me every day, for encouraging me to follow my dreams, and for making me the person I am today. Moreover, I would like to express my gratitude to my friends for being there for me whenever I needed them.

Finally, I would like to thank my Grohe colleagues, who always supported me during my internship, made me always think critically focusing on the objectives to be accomplished and helped me become a better person and a better professional.

Resumo

A atribuição de datas de entrega tem sido estudada por muitas décadas e atraiu muito interesse recentemente. As lacunas na literatura mostram que há a necessidade de se estudar as variáveis que estão associadas às datas de entrega, utilizando especificamente modelos de regressão linear.

Esta tese concentra-se em abordar essa lacuna, apresentando um modelo de regressão linear para prever uma das variáveis mais importantes relacionadas à atribuição de datas de entrega, o tempo de fluxo de produção. Para realizar esta estimativa, foram utilizados dados do software Rockwell Arena, cedidos pelo professor Aydin Teymourifar, para avaliar a precisão do modelo e perceber se o modelo desenvolvido se pode revelar uma ferramenta útil para ser aplicada em empresas de manufatura.

A análise feita na secção dos resultados, mostra que as variáveis independentes do modelo contribuíram significativamente para explicar o tempo de fluxo das produções e que o modelo tem uma precisão satisfatória.

Palavras-chave: Atribuição de Datas de Entrega, Empresas de Manufatura, Regressão, Tempo de Fluxo

Abstract

The Due date assignment has been studied for many decades and has attracted a lot of interest recently. The gap in the literature shows that there is a necessity on studying the variables that are associated with due dates, and specifically using linear regression models.

This thesis focuses on addressing this gap by presenting a linear regression model to predict one of the most important variables related to the due date assignment, the flow time of productions. To perform this estimation, data from the Rockwell Arena Software, provided by Professor Aydin Teymourifar, was used to evaluate the accuracy of the model and to understand if the model developed can be a useful tool to be applied in manufacturing companies.

The analysis made in the results section, show that the independent variables of the model contributed significantly to explain the flow time of productions, and that the model has a satisfactory accuracy.

Keywords: Due Date Assignment, Manufacturing Companies, Regression, Flow Time

Number of words: 9049

Abbreviations

ANOVA: Analysis of Variance

ANN: Artificial Neural Networks

CON: Constant Flow Allowance

DDA: Due Date Assignment

DIPPSDDA: Dynamic Integrated Process Planning, Scheduling, and Due Date Assignment

DPPW: Dynamic Processing Plus Waiting Time

DTW: Dynamic Total Work Content

IPPS: Integrated Process Planning and Scheduling

JIT: Just-In-Time

MAE: Mean Absolute Error

MAL: Mean Absolute Lateness

MSE: Mean Squared Error

MSL: Mean Squared Lateness

MTO: Make-to-order

MTS: Make-to-stock

NOP: Number of Operations

PPW: Processing Plus Wait Time

RMSE: Root Mean Squared Error

SC: Supply Chains

SLK: Slack

SWDDA: Scheduling with Due Date Assignment

TWK: Total Work Content

Index

ACKNOWLEDGMENTS	V
RESUMO.....	VII
ABSTRACT	IV
ABBREVIATIONS.....	VI
INDEX	VIII
FIGURE INDEX	X
TABLE INDEX.....	XII
1. INTRODUCTION.....	15
1.1. General Framework.....	15
1.2. Objectives and Research Methodology.....	16
1.3. Macrostructure	17
2. LITERATURE REVIEW.....	19
2.1. Just-In-Time philosophy	19
2.2. Due date assignment concepts.....	22
2.3. Due Date Assignment Models.....	24
2.4. Applications of linear regression and correlation analysis.....	29
3. PROBLEM DEFINITION AND RESEARCH QUESTIONS	31
3.1. Problem Definition.....	31
3.2. Research Questions.....	32
4. RESEARCH METHODS	35
4.1. Method Definition.....	35
4.2. Model	36
4.3. Performance measures	40
5. RESULTS.....	42
6. CONCLUSION	50
BIBLIOGRAPHY	51
APPENDIX.....	55

Figure Index

Figure 1: Arena Model.....	39
Figure 2: Assumptions of the Arena Model	40
Figure 3: Correlation Matrix	43
Figure 4: Fitted versus Residuals plot	46
Figure 5: Q-Q plot	47
Figure 6: Histogram of Residuals.....	47
Figure 7: Scale-Location plot	48
Figure 8: Performance Measures.....	48

Table Index

Table 1: Variables used to develop the model	37
Table 2: Parameters used to create the data	42
Table 3: Output of the model.....	44
Table 4: Analysis of Variance (ANOVA)	45
Table 5: Literature Review	61

1. Introduction

1.1. General Framework

The manufacturing industry plays a crucial role in various sectors of society, and the effectiveness of production processes is crucial for these businesses to stay competitive. Today, rising energy costs and raw materials are considered vital challenges worldwide. In order to maintain competitiveness, many manufacturing companies look for fast technological growth, relocate their industrial operations to less expensive locations, and must take notice of the sustainability of the production process (Edh Mirzaei et al., 2021).

Concerning one of these challenges, the United Nations provide a roadmap for a sustainable future. The goal 9 “Industry, Innovation and Infrastructure” and more specifically the target 9.4 has a direct relation to the theme of this thesis: “By 2030, upgrade infrastructure and retrofit industries to make them sustainable, with increased resource-use efficiency and greater adoption of clean and environmentally sound technologies and industrial processes, with all countries taking action in accordance with their respective capabilities”.

The central concepts of this thesis are due date, flow time, and cycle time. Flow time and cycle time are two related concepts integral to process management and often used in lean management strategies. It is essential to clarify these concepts, as flow time focuses on the average time for a unit to move through the system and comprises the time spent working on the unit as well as the time spent waiting in a queue. In contrast, the cycle time is associated with the average time between the completion of jobs (Jacobs & Chase, 2018). On the other hand, the due date concept represents the deadline by which a task or project must be completed.

In the past decades, many studies and articles have been written about Due Date Assignment (DDA), which focused on developing more efficient models to assign due dates to jobs. These studies pursued an effective DDA so that the customer expectation was satisfied and the resources allocation was efficient, "early completion of a job means to bear the costs of holding unnecessary inventories, while finishing a job late results in contractual penalty and loss of customer goodwill" (Cheng, 1988).

1.2. Objectives and Research Methodology

This dissertation aims to estimate the flow time of productions using a linear regression model with simulated data. This issue is important because providing accurate due dates to jobs is an essential element of production management in manufacturing companies. More specifically, precise flow time estimations lead to better planning and scheduling decisions, enabling more informed decisions in the production process, and allocating resources more efficiently. Also, during the production process, bottlenecks can be easily identified. In the culmination of the operation, the estimation of flow time leads to stronger customer relationships by providing more reliable due dates (Mokhtari & Kianfar, 2007).

Although much research has been done on various models to improve the DDA in manufacturing companies in the last decades, and even though there is research on flow times in manufacturing, there is a considerable research gap in studying specifically the use of multivariate regression analysis in this context. By filling this gap, this thesis intends to contribute to the research and provide significant insights of the factors that influence flow time in manufacturing organizations.

For the reasons listed above, calculating the flow time of productions accurately is necessary to determine effective due dates for jobs. This dissertation's goals are to understand how a linear regression model can be used

to predict flow time and to conduct a deep analysis of the variables that have the greatest and least significant effects. This dissertation will concentrate on two study questions:

How can we estimate the flow time of productions with a linear regression model? What are our independent and dependent variables?

In this research, we outline a conceptual method for developing a multivariate linear regression model with simulated data. Our goal is to use this model in a manufacturing company to optimize production flow time by identifying the variables that have a big significance on it. However, a major obstacle is found in the implementation of this model due to missing data.

The primary goal was to apply this model into practice at the Grohe Albergaria plant, one of many facilities owned by this German company that produces sanitary fittings. Grohe is a manufacturer with a reputation for producing high-quality goods at high quantity and having a significant workforce of over 900 people at the Albergaria plant. We selected Grohe because it is a manufacturer that regularly uses lean production techniques and has won numerous international awards, including the German Sustainability Award Design in 2021. However, we found that a portion of the data collected was missing. In particular, the missing data was the waiting time, one of the most important components of the flow time of productions. We could not successfully use our model because the company's data warehouse lacked accurate waiting times for each production job. Despite this setback, we think that manufacturing firms will find our conceptual method using simulated data to be successful.

1.3. Macrostructure

The structure of this dissertation is organized in 6 chapters. The first chapter, the introduction, presents the range of the research, where topic generalizations are made, and some items of previous research are reviewed.

Also, it is identified a research gap, along with the respective research questions. The chapter ceases with the presentation of the overall structure of this dissertation. The second one, the literature review, summarizes the current state of knowledge in this area of research. Then, in the third chapter, a profound description of the problem is made, and consequently, the respective research question. The methodology is the next section to be presented, where the strategy of this dissertation and the procedure that is used for data collection are presented. The fifth chapter focuses on displaying the experimental results based on the methods used. In this section, an analysis of the data collected is made and following it, the model analysis is disclosed. Finally, the sixth chapter summarizes the key findings, the limitations and the implications of the research. Recommendations will then be given for future research, and for the company to record data in a better way to assign more effectively the due dates.

2. Literature review

2.1. Just-In-Time philosophy

The evolution of supply chain (SC) throughout the globe is within reach due to a continuous contest between industries. This fight for achieving competitive advantage and sustainability throughout the SC enhances manufacturing companies to reduce waste and to improve its efficiency by producing only what is necessary at the time it is necessary, the Just-In-Time (JIT) production (Jacobs & Chase, 2018).

Due to the growing interest in JIT production in the industry, earliness-tardiness completion time of jobs has received a lot of attention. Thus, the need for improving SC networks were related with completion time penalty costs linked to lead times, tardiness and earliness deliveries (Z.-L. Chen, 1996; Seidmann & Smith, 1981).

Effective production management involves balancing the competing demands of cost, time, and quality. One key area of focus is minimizing the cost of early and tardy execution penalties. By scheduling the completion of each task as close to its deadline as possible, it is possible to reduce these costs, and ultimately improve overall production efficiency.

In case a job is finalized earlier and kept in inventory, it can arise a fee known as an earliness penalty. The repercussion of this can be inventory storage fees, degradation, insurance, and others. On the contrary, depending on the lead time and the respective due date, there may be impossible to minimize the total tardiness of a job completion, leading to a lateness penalty which can damage costumers' opinions, lead to reputational harm, or even indemnitee the manufacturing company with contract fines (Z.-L. Chen, 1996).

Rather than sticking to a rigid plan, agile SC use techniques that allow them to quickly adapt to changes in customer demand. Because of this, agile SC can be

responsive and flexible to customer needs, which is important in today's fast-paced business environment. Additionally, these SC also look to minimize the risk of supply shortages or disruptions. This is done by being able to anticipate and respond to changing, diverse, and unpredictable customer demands on the front end, while also minimizing risks on the back end (Jacobs & Chase, 2018).

Thus, the later the due dates are set, the more likely the item will be finalized or on time delivered. To maintain a positive image among customers, many businesses are willing to accept reasonable holding costs in exchange for meeting established deadlines. As a result, the decision maker must weigh the losses caused by holding costs against the benefits of completing orders on time (V. S. Gordon & Strusevich, 2009).

As interest to the research, two production strategies are presented in which revenue is accomplished with efficient scheduling: Make-to-stock (MTS), Make-to-order (MTO), and other situations in which efficient scheduling leads to a company becoming the first mover with an increased market share: new product markets. Moreover, it is perceived that reducing lead times becomes more important than ever in MTO and more inventory presence is important for the customer in MTS (Z. L. Chen & Hall, 2008).

The growing demand for on-demand products and services with increased flexibility, combined with the accelerating socioeconomic environment in the age of Industry 4.0, has compelled manufacturers to rethink their current strategies, which primarily prioritize time, cost, and quality. By implementing practices that improve their ability to respond to changes in customer demand quickly, companies give rise up to a new a critical competitive advantage: responsiveness (Jacobs & Chase, 2018).

Operations managers can provide a variety of high-quality products while also lowering expenses. This is made possible by mass customization, which enables the speedy and low-cost manufacture of goods and services that satisfy the unique requirements of clients. This strategy not only broadens the selection

but also enables products to be created precisely to the customer's specifications at a fair price. But putting mass customization into practice can be challenging and calls for advanced operational abilities, like agile processes, constrained product lines, and modular design. The business's various departments, including sales, design, production, SC, and logistics, must work closely together (Heizer et al., 2017).

Parallel to other essential lean methods such as automation, kaizen/continuous improvement, total productive maintenance, and value stream mapping, the literature presents JIT (Garza-Reyes et al., 2018), a component of lean manufacturing, and its relation with agile manufacturing. This methodology places a strong emphasis on quick customer response. Interconnected it is being argued that combines the primary tactical elements of JIT-production and JIT-purchasing to reduce waste and maximize resource utilization throughout the SC. When combined with JIT-production, JIT-purchasing improves a company's manufacturing agility. Additionally, organizations that transform into agile manufacturers can expect improved operational and firm performance (Inman et al., 2011).

Agility and JIT have a close relationship to manufacturing and production. By creating goods only as needed, JIT allows businesses to lower their inventory costs and boost productivity. As a result, they can quickly adjust to fluctuations in demand and better meet client expectations.

A crucial component of JIT is DDA, which ensures that products are manufactured and delivered to clients on schedule. Companies can give due dates that consider the present circumstances and reduce the risk of delays by using precise and up-to-date information on the status of tasks and the current shop load.

In conclusion, the JIT philosophy, DDA, and agility are all connected. Businesses that use KIT philosophy must be flexible and adapt to changes in the marketplace and environment, and by using accurate DDA methods, they can

reduce the risk of delays while boosting productivity and customer satisfaction.

2.2. Due date assignment concepts

The due dates assignment strategies for companies go back long before the twenty-first century. The term "due window" was initially used in scientific research in this field as a generalization of the term "traditional due date" and to describe a period rather than a single precise moment. Most due date issues at this point are connected to the placement of due windows, incurring in penalties for having completed the scheduled time earlier, earliness penalties, or after the planned time, tardiness penalties (Janiak et al., 2015).

Concerning the DDA, the literature relates it to the scheduling concept and how scheduling operations may affect manufacturing operations. Specifically, it is mentioned when manufacturing companies create schedules that match resources to client demand and focus on producing items JIT with short setup periods, little processing times, and high percentage utilization these manufacturers can reduce expenses and meet due dates by efficiently organizing their work (Jacobs & Chase, 2018).

DDA plays a crucial role in the production planning strategy of manufacturing companies. The approach can take the form of forward scheduling or backward scheduling. In forward scheduling, job demands are first identified, and a schedule is created to complete the job within the earliest possible timeframe, aligning with customer orders and planned for the earliest delivery windows. In contrast, backward scheduling starts with the final due date and works backwards, scheduling each task in reverse order and determining the start time by subtracting the time required for each task. To find a balance between meeting production capacity limitations and fulfilling customer expectations, companies often use a combination of both forward and backward scheduling methods (Heizer et al., 2017).

In the field of production planning and scheduling, literature has identified

the concept of "loading" as a key aspect of DDA. Loading refers to the process of allocating tasks to processes. There are two main types of loading: finite loading and infinite loading. Finite loading allocates capacity to workstations based on their ability to handle the work, which in theory allows all jobs to be completed. However, this method may result in postponing due dates to a later time, as the workstations are only loaded with what they can handle, even if there is more work available. On the other hand, infinite loading assigns tasks to workstations without regard to capacity constraints, which allows for adhering to deadlines, but may result in overloading workstations. Naturally, either the capacity or the timetable must be changed when the workload surpasses the available resources (Heizer et al., 2017).

DDA in manufacturing companies also involves managing situations where jobs are behind schedule or have had their completion dates pushed forward. To meet the target due date, managers may need to accelerate the completion of remaining tasks by allocating additional resources. Another concept involved in the literature is "project crashing". Related to it, the "crash time", which is referred as the minimum amount of time that is required to complete the task, in order not to have a "crash cost". This concept arose with the necessity of understanding how a project could be finished as fast as possible. It is common to shorten a job by increasing resources such as tools or personnel, therefore, it is logical that the crash cost of an activity would be higher than its average cost (Heizer et al., 2017).

Meeting due dates is closely connected to reducing setup costs, minimizing inventory levels during job completion, and maximizing machine and worker utilization. However, achieving all these goals simultaneously can be challenging. For example, even if a high percentage of due dates are met, missing the deadline for a crucial project or important client can still result in a significant setback in the overall schedule. It is important to adopt a holistic approach, known as a "systems perspective," to align the goals of the value stream with the

organization's operations strategy (Heizer et al., 2017).

One of the key considerations when assigning due dates is the impact of variability in production processes. Variations in processing times, changes in material supplies, machine breakdowns, shifts in consumer demand, human errors, and absenteeism can all contribute to unpredictability in the manufacturing system. To account for this unpredictability, production plans must incorporate contingencies such as lead times or safety due dates. However, a major challenge in setting due dates is understanding how the various uncertainties in process times interact with one another through the product structure, and how these interactions ultimately affect the deadline for the finished product (Song et al., 2002).

2.3. Due Date Assignment Models

The industry's JIT philosophy, which states that tasks should be finished as near to their due dates as possible, gave rise to scheduling research using irregular performance indicators. Due date determination approaches for issues with non-regular performance measurements have been developed because of the growing interest in the pursuit of optimal decisions in JIT systems. As a result, a substantial body of literature addressing DDA models and scheduling issues has emerged during the past thirty years (V. S. Gordon et al., 2002).

Scheduling models can be broadly categorized into static and dynamic models. The term "dynamic scheduling problem" refers to a scheduling issue when stochastic events, such as the arrival of new jobs, equipment failures, and order cancellations, occur over time. Most of the scheduling research emphasized static job shop scheduling. All scheduling criteria must be static in a static job shop schedule, and all work information must be known and available at time $t = 0$. Even though improving static job shop scheduling is a well-known issue in the literature, dealing with dynamic occurrences is necessary to boost productivity and machine balancing rates in a real manufacturing system.

Because suggested plans and schedules must be modified to account for interruptions, the most recent production scheduling models are inherently dynamic and more challenging to solve than static models (Erden et al., 2019).

Stochastic models are widely used for DDA in production planning and scheduling. Among the most well-known models are: Constant Flow Allowance (CON), which consider the constant flow of work through the production process, and models that consider processing times, such as Slack (SLK), Total Work Content (TWK), Processing plus Wait (PPW), and Number of Operations (NOP).

In scheduling literature, " d_j " is used to refer to the due date of job " j ". There are several models for assigning due dates, including the simple model in which all jobs have the same due date, referred to as the CON model, which is often used in situations where tasks need to be completed at the same time, such as in an assembly system or a shop where multiple jobs make up a single customer's order (V. S. Gordon et al., 2002).

The SLK model of assigning due dates involves using a common slack value, referred to as " q ", which is added to the combination of release dates and the processing times for each job, resulting in the equation $d_j = r_j + p_j + q ; j = 1, \dots, n..$. When all jobs start at the same time, at time zero, the equation simplifies to $d_j = p_j + q$. This model reflects the scenario where the facility oversees setting due dates, and the decision maker calculates the flow times for each job. The common slack factor is used when all jobs are intended for the same customer or when it is important to demonstrate consistent decision making (V. Gordon et al., 2002). Thus, this model is used in single machine systems to estimate realistic delivery dates and efficiently arrange the orders in order to avoid late payment penalties (Shabtay & Steiner, 2007).

The TWK model for determining due dates, involves setting the due date as the release date plus a multiple of the job processing time, represented by the equation $d_j = r_j + kp_j$. The constant " k " is used to define the due date for each job,

and it must be greater than 0. If it is assumed that all jobs are available at the same time, the due date is set as $d_j = kp_j$ (V. Gordon et al., 2002).

The CON, SLK, and TWK due dates are combined into one model, the PPW model, which due dates are linear functions of the job processing times. The due dates in the PPW model are determined by the formula $d_j = r_j + kp_j + q$, where $k \geq 0$, is a common multiplier and q is a slack allowance that may be negative. Finding the ideal values for k and q is what makes an assignment have the best due date. If it is assumed that all jobs are available at the same time, represented by $r_j = 0; j = 1, \dots, n..$ and in case the jobs are processed on a single machine without interruption or machine idle times, the schedule is defined by a sequence of jobs. The PPW due dates then are calculated as $d_j = kp_j + q, j = 1, \dots, n$ (V. Gordon et al., 2002).

In the NOP model, the due date is based on the number of operations required to complete the job, represented by the equation $d_j = kM_j$, where " k " is the constant that needs to be determined and " M_j " is the number of operations required for job " j ". This model is typically used in multi-machine environments, but it can also be used in single-machine situations (V. Gordon et al., 2002).

Additionally, several dynamic DDA models have been discussed in the literature, including the Dynamic Total Work Content (DTWK) and Dynamic Processing Plus Waiting Time (DPPW) models.

Regarding TWK, it is a model that assigns the same level of tightness to all jobs based solely on the total processing time of a job. However, this approach doesn't consider the current shop load, which means that if two jobs have the same amount of work, they will be given the same allowance regardless of whether the shop is busy or not. To improve performance in terms of meeting job due dates as closely as possible, a modified form of the TWK model can be used that employs a dynamic due date allowance factor, the DTWK. This would mean that when the shop is busy, a longer flow time allowance would be assigned to an incoming job and when the shop is not busy, a shorter allowance would be given.

Similarly, the PPW due date model can also be improved by using a dynamic due date allowance factor based on feedback information, the DPPW. This will enable a more precise estimation of flow time (Cheng & Jiang, 1998).

Furthermore, it is challenging to predict the completion times for jobs in a dynamic job shop environment, where jobs are continually entering and leaving the production system in a random fashion, as the completion time can be affected by a variety of factors related to the job and shop characteristics. Researchers have used multiple performance measures when studying the scheduling characteristics of dynamic job shops. Hence, the improved due date models can adjust their flow time predictions based on the current shop workload using feedback information. According to simulation results, these dynamic models are more effective at avoiding missed due dates compared to their previous, static versions, and minimize Mean Absolute Lateness (MAL) and Mean Squared Lateness (MSL) for DDA (Cheng & Jiang, 1998).

Connected to it is the JIT philosophy, which holds that both being early and being late are undesirable, therefore it's more suitable to consider both tardiness and earliness as "missed due date". The measure of MAL is a good linear measure for missed due dates. However, when the precision in meeting target due dates is more important, non-linear measures of missed due dates are more appropriate, such as MSL which is a quadratic function of missed due dates and thus imposes a higher penalty for larger deviations from the designated due date (Alpay & Yüzügüllü, 2009).

Researchers have been interested in studying ways to improve the efficiency of manufacturing systems by integrating different functions, such as process planning and scheduling. This can be divided into two main categories: Integrated Process Planning and Scheduling (IPPS) and Scheduling with Due Date Assignment (SWDDA). IPPS aims to improve scheduling performance by optimizing factors such as the overall time needed to complete a job, earliness and tardiness of jobs, and the overall equipment utilization level. Many studies

have been conducted on IPPS to find more effective ways to plan and schedule manufacturing processes. Similarly, there is also a lot of research on SWDDA, which focuses on ways to improve the scheduling and assignment of deadlines in manufacturing to make the process more responsive (Erden et al., 2019).

Moreover, literature talks about Dynamic Integrated Process Planning, Scheduling, and Due Date Assignment (DIPPSDDA). This model of managing production processes in a manufacturing or production environment. It involves the integration of different planning and scheduling activities, such as process planning, material planning, and production scheduling, to optimize the use of resources and improve overall production efficiency. The goal of it is to minimize lead times and improve delivery performance by taking into account factors such as production capacity, materials availability, and due date requirements (Erden et al., 2019).

Furthermore, some experiments have been tested to solve DIPPSDA (Demir & Erden, 2020; Erden et al., 2019). It is shown that Genetic Algorithms/Tabu Algorithms and Tabu Algorithms show higher reliability to solve DIPPSDDA and that not only GA achieve good performances, but also Ant Colony Optimization algorithms can solve DIPPSDDA problems.

Artificial Neural Networks (ANN) are also discussed in the literature as a DDA model (Teymourifar & Ozturk, 2018). It was proposed two ANN models that were compared with previously established and successful models such as TWK, PPW, DTWK and DPPW. These models created were successful mostly by considering multiple factors related to the environment in order to estimate the waiting time for a job, which is a more effective approach than simply estimating the direct finish time.

Proceeding, the literature also talks about variables that are part of DDA models. More specifically, in a dynamic job shop, where jobs of various types are constantly entering and leaving the production system in a random way, the amount of time it takes to complete a job is influenced by many factors. In order

to predict the processing time accurately, will depend on the characteristics of the production system, such as the dispatching rule used and the rate of utilization of the shop (Chang, 1994).

Also, behavioral phenomenon that affects the productivity of the scheduling system is ignored in the scheduling literature. In the workplace, human multitasking—in which a person moves from one activity to another often—occurs as often and typically causes a decrease of performance. The ability to deliver at the agreed-upon due window is compromised because the processing of a chosen task is interrupted by many other tasks that are available and yet unfinished. Research into this phenomenon would increase scheduling's effectiveness because it has an impact on total processing time, lateness penalties, and scheduling (Zhao et al., 2017).

2.4. Applications of linear regression and correlation analysis

Regression analysis is a statistical methodology due to its ability to establish a link between variables using simple procedures and can be utilized to module the future relationship for them. Thus, there is a wide range of applications that have an impact on society, including agriculture and environmental sciences, government, history, manufacturing, healthcare, and many others. More into the point, linear regression is known as important technique for estimating in manufacturing firms since it enables informed decision making based on data-driven forecasts (Chatterjee & Hadi, 2012).

Studies about the variables that affected the performance of manufacturing companies have started many decades ago. Using a linear regression model, research using a sample of firms in 100 manufacturing industries was made to analyze different aspects of company dynamics. The correlations analysis made revealed that as a company gets older, its growth, variations in growth, and the likelihood of failure all diminish. Furthermore, even after removing low-growth firms from the sample, the study discovers that the rate of a firm's growth drops

at a decreasing rate with the firm's size (Evans, 1987).

More recently, a study on machine-tool selection employed a linear regression model to identify the best machine to meet the needs of a manufacturing organization. Choosing ineffective equipment can have a detrimental impact on the company's performance. As a result, this is a critical process that may present some difficulties for the decision-maker. The aim of this research was to create a cost model based on machine attributes. This model's applicability proved to be useful for decision-makers in the early phases of a selection process because they provide a quick tool for generating cost approximations and eliminating the need for huge and costly research (Ciurana et al., 2008).

Another research used a linear regression model to calculate tool wear (a gradual failure of a constant use of the tool). By estimating it continuously, the machine could better perform and machine components were more accurate, by having establish a correlation between the independent and dependent variables. (Karuppusamy et al., 2015).

On another area concerning manufacturing companies, a study examined the effect of numerous financial components on profitability ratios, such as return on assets and return on equity. Furthermore, it was possible to examine the clear effects of explanatory variables on profitability indicators. For this study, a correlation analysis was not enough, having been also done a regression analysis that proved that the variables had implication on financial policies for manufacturing companies in Ghana (Amponsah-Kwatiah & Asiamah, 2020).

3. Problem Definition and Research Questions

3.1. Problem Definition

In manufacturing companies, it is important to meet customer expectations by completing orders not earlier, not later, but on time. To achieve this, due dates are assigned to orders based on estimates of the time required to complete the different steps of the manufacturing process.

However, accurate estimation of flow time can be challenging due to several variables. The variables in which flow time is calculated is with the sum of the processing time with the waiting time. The production process is impacted by these variables. If other products in the production line take longer to process, it may create congestion and delays in the production line, increasing the flow time of all products. This can have a succeeding effect on the production process, causing additional delays, specifically in manufacturing companies that use FIFO methods.

Another key independent element that might influence product flow time in manufacturing companies is the capacity designated to product a good. If the assigned capacity is not enough, it might cause delays and congestion in the production line, increasing product flow time. If the given capacity is too large, the manufacturing line may be underutilized, resulting in decreased efficiency.

The average arrival rate of products to the production line is another variable with a relevant importance for this study. The quantity of production materials arriving at the production line can influence the time it takes to complete each good. On one side, a high arrival rate can cause delays, but if the arrival rate is too low, the production line may be underutilized.

Another factor to consider is overall equipment effectiveness, which measures the efficiency of the machines. This is important because malfunctioning machines can cause delays in the production process and have a negative effect

on the result.

Because many products are assembled from tiny pieces, it is critical that all components required for manufacturing arrive on time and in good condition. If components are delayed or damaged, it can cause postponements in the production process.

Finally, there is a variable that is difficult to measure its impact, which is the operator's behavior and mood, as this can affect the performance of the machines and the quality of the output. If operators are tired, stressed, or not motivated, it can lead to mistakes and lower quality products.

To summarize, to deliver orders on time in manufacturing organizations, this thesis will focus on the variables that affect flow time estimations. In case companies don't look up these factors that affect the variability of flow times, it will impact negatively on meeting customer expectations, planning production, managing inventory, costs, and on competitive advantage.

3.2. Research Questions

In this study, we investigate the thesis theme by posing the following research questions: "How can we estimate the flow time of productions using a regression model? What are the independent and dependent variables involved?"

The objective of the research is to analyze the current methods used for assigning due dates in manufacturing companies, and to propose a new model that enhances the efficiency and effectiveness of the process. The significance of this research lies in the fact that proper assignment of due dates is crucial for the smooth operation and success of manufacturing companies.

This thesis aims to explore the variables that affect the flow time of production, with the goal of providing useful insights for manufacturing companies and researchers interested in the field of production management.

In an early stage, a model was developed, in which the goal was to study the variables that affect the flow time of production of finished goods on Grohe's

Albergaria Plant assembly lines, and consequently influence the efficiency of the allocation of due dates to jobs. With that intention, it was studied a multivariate linear regression model, and its respective dependent and independent variables. It was observed from the data available, that certain variables such as the waiting time, lacked reliability, being challenging to evaluate whether they were accurate or inaccurate. In other words, judgments about the model created could not be made based on this data, putting in view the lack of accuracy.

Even so, this thesis fills a major gap in existing research by investigating the variables that have the greatest impact on the flow time of production using a linear regression model to generate reliable estimations. The goal of this study is to create a multivariate regression model that can estimate the dependent variable, the flow time of productions, by taking several parameters into account at the same time. This thesis intends to contribute to a more thorough understanding of production flow time and provide significant insights for enhancing operational efficiency in manufacturing environments by taking this approach. In addition, the production of goods daily in manufacturing companies is interfered by many factors that can be either controllable and easy to riddle, but also factors that are unexpected and challenging to solve.

On one side, the regression model presented in this thesis has got four independent variables that were obtained by a simulated model: the total processing time of other process products at the factory at the time of order of product; the assigned capacity to produce the respective product; the processing time; the flow time of the production.

Even though this model does not consider specific variables such the OEE, the most important variables to the model are presented with unequivocal accuracy, with the goal to support manufacturing companies to anticipate decisions that regard the allocation of due dates to jobs.

Manufacturing companies may maintain smooth and effective operations, increasing productivity and lowering costs, by keeping an eye on production

indicators and being proactive about delays and orders that finish before the estimated time.

4. Research Methods

4.1. Method Definition

Regression analysis is a statistical methodology that offers researchers a robust tool to analyse and measure the correlations between variables. This strategy asks researchers to build a model that connects a dependent variable to one or more independent variables. By developing these models, researchers can acquire significant insight into the elements that influence the outcome of interest. Although regression analysis is conceptually simple, it is a frequently utilized and versatile method in various study disciplines.

Regarding the regression model equation $Y = f(X_1, X_2, \dots, X_i) + \varepsilon_1$, we can observe a relationship between the dependent variable Y and the independent variables X_1, X_2, \dots, X_i . Also, in this formula, the value represented by ε corresponds the error in of the prediction of Y given the values of X_i variables (Chatterjee & Hadi, 2012).

In this statistical method, the relationship between the dependent and independent variables is directly proportional, which is then observable by the best fit straight line in which the differences between the predicted and the actual values are the minimum possible. This type of analysis is practical way to determine the strength of the predictor variables, and an accurate way to understand their effects and significance on the response variables.

In this way, a linear regression model can be represented by $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon_1$, where β_0 is the intercept or the value of Y when all the predictors are zero and from β_1 to β_p the coefficients to be determined (estimated) from the data collected. In other words, Y can be approximated as a linear function of X_1, X_2, \dots, X_p and the constant term ε represents the discrepancy in that approximation. It should be noted that e does not provide any additional information for determining Y beyond what is already captured by the

independent variables. Therefore, the validity of the linear regression model depends on the range and quality of the independent variable data used (Chatterjee & Hadi, 2012).

To investigate a deeper relationship between the variables of the study, a correlation matrix was developed, and the respective analysis was made.

Specifically, it will be presented an investigation of the relationship between the variables of the model selected. This method is predicated on the notion that the variables have a linear relationship, being able to assess the strength and direction of the link between the variables, similar to measures of association for binary variables. A correlation coefficient, which varies from -1 to +1, will result from this investigation. Whereas a coefficient of +1 denotes the presence of a perfect positive linear relationship and a value of -1 represents the presence of a perfect negative linear relationship, a coefficient of zero refers to the absence of any relationship at all (Gogtay & Thatte, 2017).

4.2. Model

During this investigation, it was developed a multivariate regression model, that focused on understanding the relationship between the variables in the table below. This study had the objective not only to approximate de equation of the model presented to the calculation of the flow time of production that is made in manufacturing companies, but also to incorporate the variables that could have the most important significance to the response variable.

Variable	Description
λ_i	Mean arrival rate when product i arrives to line
Q_i	Total processing time of other products at the factory at the arrival time of product i
Ca_i	Assigned capacity to produce product i
R_i	Release time of product i
P_i	Processing time of product i
D_i	Due date of product i
W_i	Waiting time of product i
C_i	Completion time of product i
F_i	Flow time of product i
\hat{F}_i	Estimated flow time of production i

Table 1: Variables used to develop the model

As a starting point of the investigation, P_i , the completion time of product i , was the first variable to be studied. From it, we can understand that it was obtained by the sum of the following variables: release time of product i , R_i , the processing time of project i , P_i , and the waiting time for project i , W_i . The Equation 1 illustrates the composition of the Completion time of product i .

$$C_i = R_i + P_i + W_i \quad (1)$$

Another variable considered for the model is the total time taken for product i to go through the entire process from start to finish, the production flow time for product i , F_i . This variable is obtained by summing both the processing time, P_i , and the time that product i spends waiting before it can start processing, W_i . Flow time is expressed in Equation 2:

$$F_i = P_i + W_i \quad (2)$$

Moreover, by substituting the equation 2 into the equation 1 we can obtain:

$$C_i = R_i + F_i \quad (3)$$

On top of that, related to this last expression of the completion time of project i , C_i , the DDA variable of the product i , D_i , can be expressed not only by integrating on it the release time, R_i , at which the product is released to for processing, but also, the estimated flow time, \hat{F}_i , which is the predicted time will take for the product to go through the entire production process. Therefore, the variable D_i can be expressed by equation 4:

$$D_i = R_i + \hat{F}_i \quad (4)$$

As a result of the importance of the variables above explained, a multivariate linear regression model is developed to estimate the flow time of productions, \hat{F}_i , based on several predictor variables.

For this model, the predictor variables chosen are: the mean arrival rate when product i arrives to line, λ_i ; the assigned capacity to produce product i , Ca_i ; the processing time of product i , P_i ; the total processing time of other products at the factory at the time of production of product i , Q_i .

The multivariate regression model can be presented with the following equation:

$$\hat{F}_i = \beta_0 + \beta_1 \lambda_i + \beta_2 Ca_i + \beta_3 P_i + \beta_4 Q_i + \text{Epsilon} \quad (5)$$

In the equation above, beta, β , is the interception and coefficient of the regression. Epsilon is error term, i.e., the difference between the actual value and the observed response value. It's also important to add that this model uses hours as its unit of time.

Additionally, to verify the validity of the model presented in equation 5, the simulation method is employed. Since the thesis is not centred on simulation, we have not included the simulation's details in the thesis. The simulation was run

by the thesis supervisor. However, interested researchers can get more information by getting in touch with us.

The Rockwell Arena 14.0 was the software chosen to collect the data, which uses an entity-based methodology. The flowchart presented in the Figure 1 represents the assumptions that were used to perform it. The goal of the simulation performed was to imitate a system, which could be a real or conceptual. It was chosen to imitate a conceptual model. In this way, the model recorded the variables used for each product, in our case, the arrival rate, the capacity, and the total processing time of other products at the arrival time. After the processing time is in the line, the completion time is then recorded.

In Figures 1 and 2, the Arena model and its assumptions are presented.

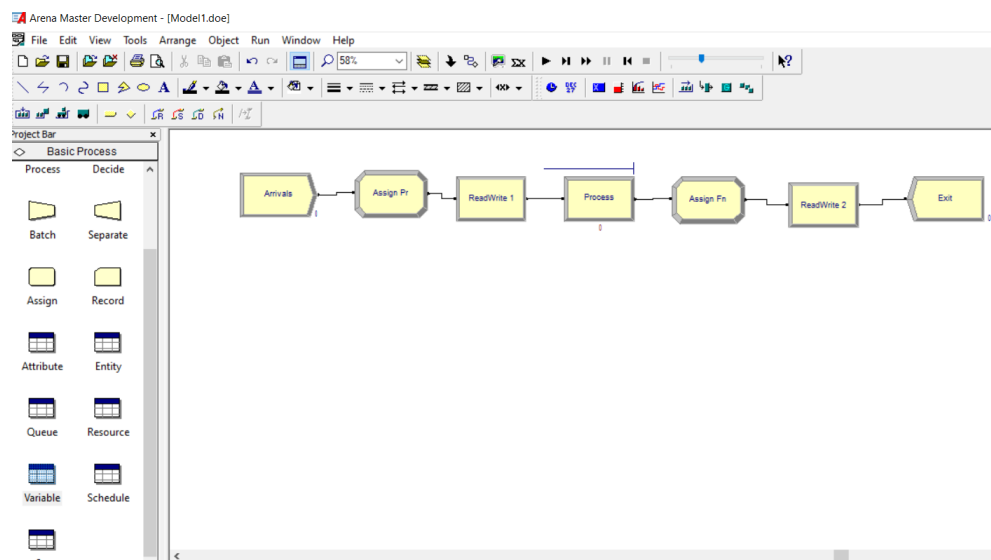


Figure 1: Arena Model

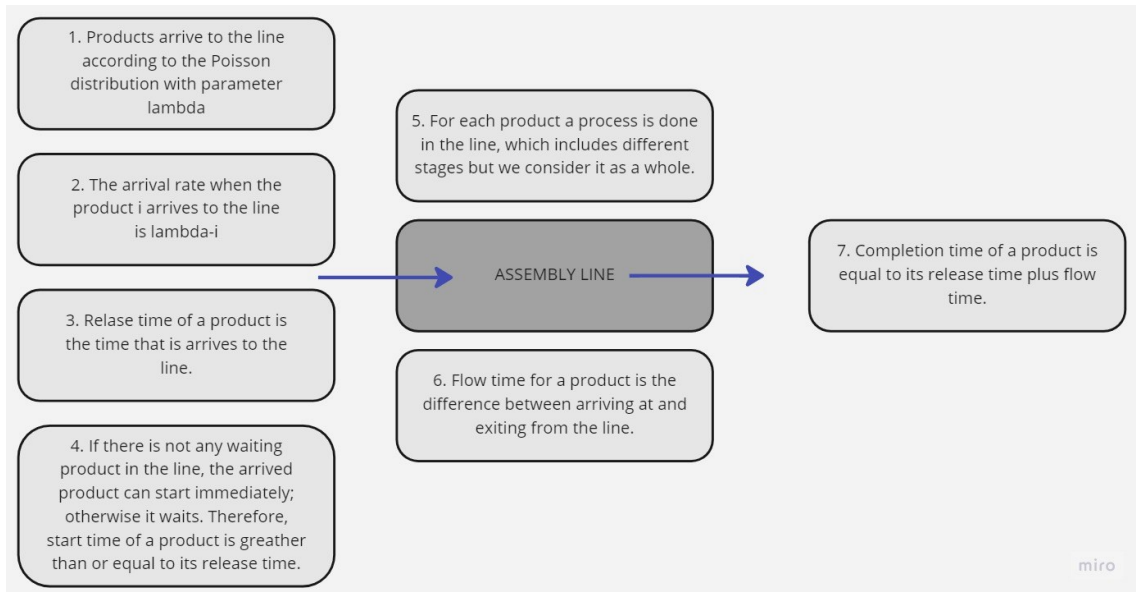


Figure 2: Assumptions of the Arena Model

4.3. Performance measures

There are many performance measures that are covered by different linear regression models, which are useful to present to the public. The most known metrics will be used to assess the model: Mean Squared Error (*MSE*; R-squared; Root Mean Squared Error (*RMSE*); Mean Absolute Error (*MAE*); Adjusted R-squared and the F-statistic.

Starting with the first one, *MSE* is a relative measure used to understand how well the model fits the dependent variable, in our case, the flow time of productions. The average of the squared discrepancies between the predicted and observed values are quantified. A model with no errors has an *MSE* of 0, and as model error grows, so does the *MSE* value. The *MSE* in regression estimates the average squared residual, and as the data points get closer to the regression line, the model has less error, resulting in a lower *MSE*. The formula is presented below, where *n* is the number of observations of the model:

$$MSE = \frac{\sum(y_i - \hat{y}_i)^2}{n} \quad (6)$$

Secondly, the *RMSE* measure has a relation with the previous mentioned measure being the square root of the *MSE*. It measures the average difference between the predicted and the actual values.

R-Square is a statistical measure used in regression models to demonstrate how well the data fits the model. It calculates the fraction of the variance in the dependent variable that can be explained by the independent variable. This measure is a number that ranges from 0 to 1, with higher values suggesting a better match. A higher R-Squared indicates that the independent variable can explain more of the variance in the dependent variable. Also, Adjusted R-square, another measure of how a model fits the data, differs from R-square because it penalizes the R-square value by including variables that do not significantly improve the fit, and it increases in case the addition of an independent variable improves the model regarding what would be expected from it.

Last, but not least important, the *MAE* is a popular regression analysis statistic that evaluates the average magnitude of the residuals between the actual and projected values of the dependent variable. It is derived by averaging the absolute differences between the anticipated and actual values of the dependent variable for each data point. The formula can be found below:

$$MAE = \frac{\sum |y_i - \hat{y}_i|}{n} \quad (7)$$

5. Results

In this section, the parameters that were used to collect the data are presented, and an analysis of the data is performed, using a correlation analysis of all the variables of the model. The output of it and its respective equation is then shown, with a deep analysis of the coefficients of the constraints. At the end, performance measures are used to assess the quality of the multivariate regression model.

Starting with the simulated model, the parameters used to collect 5000 products are described in the table below, being the parameter P according to Gamma distribution.

The dataset contains five numerical variables, including four predictor variables and one dependent variable: Mean arrival rate when product i arrives to line, λ_i , the assigned capacity to produce product i , Ca_i , the processing time of product i , P_i , the total processing time of other products at the factory at the arrival time of product i , Q_i . The dependent variable is the estimated flow time of productions, \hat{F}_i .

Parameter	λ	Ca	P
1	1	1	1
2	1	2	1
3	2	2	1
4	2	1	2
5	1	1	2

Table 2: Parameters used to create the data

To understand better the relation between the multiple variables of the model, a correlation matrix was performed.

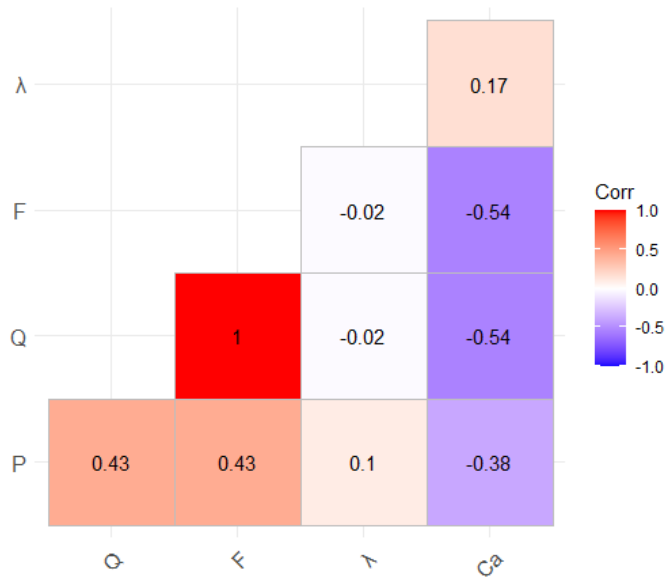


Figure 3: Correlation Matrix

The correlation matrix provides insight into the strength and direction of the linear association between each pair of variables.

Starting with the mean arrival rate when product i arrives to line, λ_i , this variable has a low positive relationship with the assigned capacity to produce product i ($Ca_i = 0.17$) and also with the processing time of product i ($P_i = 0.10$). On other side, it has a small negative correlation with total processing time of other products at the factory at the arrival time of product i ($Q_i = -0.02$) and the estimated flow time of production ($\hat{F}_i = -0.02$).

Regarding the assigned capacity to produce product i (Ca_i), this variable has moderate negative correlation both with the total processing time of other products at the factory at the arrival time of product i ($Q_i = -0.54$), the dependent the estimated flow time of productions ($\hat{F}_i = -0.54$) and the processing time of product i ($P_i = -0.38$).

Another variable, the processing time of product i , P_i , has a positive correlation with both the total processing time of other products at the factory at the arrival time of product i ($Q_i = 0.43$) and the estimated flow time of

production ($\hat{F}_i = 0.43$).

Finally, the strongest correlation (1.00) observed is between the total processing time of other products at the factory at the arrival time of product i (Q_i) and the estimated flow time of production, \hat{F}_i .

The linear regression model developed had the following output:

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	2,954875243	0,128450548	23,00399097	2,1826E-111	2,703055776	3,20669471
λ	0,520918078	0,057482409	9,06221721	1,79949E-19	0,40822732	0,633608837
Ca	-1,868766405	0,069128462	-27,03324134	2,4909E-150	-2,00428854	-1,733244269
P	1,038046996	0,012964078	80,07102278	0	1,012631711	1,06346228
Q	1,000549836	4,52192E-05	22126,65527	0	1,000461186	1,000638485

Table 3: Output of the model

By analysing the table above, it is possible to interpret that the equation of the model can be represented as:

$$\hat{F}_i = 2,9548 + 0,5209\lambda_i - 1,8687Ca_i + 1,0380P_i + 1,0005Q_i + \varepsilon \quad (8)$$

The intercept term of 2.9548 represents the expected value of the dependent variable, \hat{F}_i , when all independent variables ($Q_i; P_i; Ca_i; \lambda_i$) are zero.

Regarding the coefficients of the independent variables, they represent the change in the dependent variable for a unit increase, holding all other variables constant. Consecutively, we have that a increase by one unit in the independent variable Ca_i and all the other independent variables remain constant, the dependent variable would to decrease by 1.8687 units. On the other hand, a increase by one unit in only one of the independent variables ($\lambda_i; P_i; Q_i$) and all the other independent variables remain constant, the dependent variable would to increase by 0.5209, 1.0380 and 1.0005 units, respectively.

The second column, Standard Error is directly related to the third column, t Stat. The t-statistic quantifies the statistical significance of a regression

coefficient. It is obtained by dividing the estimated coefficient by its standard error, which is a measure of how much the coefficient fluctuates among data. In this case, the t value is the distant to zero, in which we can interpret as having great evidence against the null hypothesis.

Regarding the p-value, in the fourth column, we can interpret that as the p-value of every independent variable is lower than 0.05, there is a statistically significance between each predictor variable and the response variable. Furthermore, we can reject the null hypothesis and conclude that it exist a relationship between these variables.

Int the last two columns, we are able to see the value of the confidence interval of each variable. The variable λ_i has a 95% confidence interval between [0.40822732 ; 0.633608837]. This means that we are 95% confident that the true value of the coefficeint λ_i falls within this range.

Furthermore, an Analysis of Variance (ANOVA) used to evaluate the significance of the regression model.

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	4	2817843015	704460753,7	188774680,6	0
Residual	4995	18640,11346	3,731754446		
Total	4999	2817861655			

Table 4: Analysis of Variance (ANOVA)

By interpreting the table above, is it possible to see that Regression df is the number of independent variables in our regression model. Moreover, the Residual df value is then calculated by the subtraction of the number of the four independent variables from the total number of observations (5000) minus one.

Then, the Sum of Squares (SS), represents the squared differences between the observed dependent variable and its mean, being a measure of the total of variability of the data set. By looking at the second column, because sum of

squares is 2817843015 and the Total SS is 2817861655, we can conclude that the regression model explains 99.9993% of all variability of the dataset. Moreover, the information given shows that the regression model has a high F-statistic of 188774680.6 and a significance level of 0, which means that the model is very significant and there is strong evidence to suggest that the independent variables are closely related to the dependent variable.

The fitted versus residuals plot is a graphical technique that helps us identify deviations from the assumptions of linear regression. By examining the plot presented below, we observe that linearity appears to be satisfactory, as the red line is in proximity to the dashed line. Moreover, we can determine that the homoscedasticity assumption is met since the variability of the dependent variable appears to be consistent across the data. Finally, based on the plot, we can identify two outliers, which are data points 3793 and 3989.

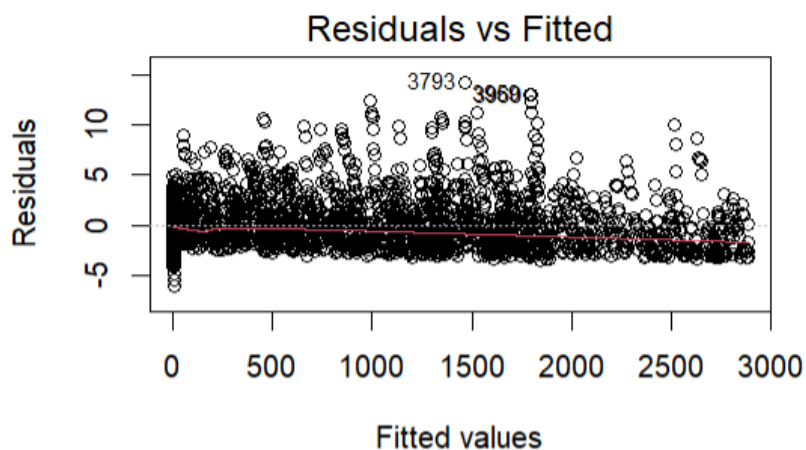


Figure 4: Fitted versus Residuals plot

In the figures below, the Q-Q Plot and Histogram of the residuals of the model can be observed. From the first, it is possible to visually evaluate how the distribution is compared to the normal distribution. Regarding the second, the histogram tells us also whether there is a normal distribution or the opposite. From the figure 5, it is possible to observe that the actual values in the upper end

of the line are deviated from the reference line. The fact that is diverging in the right end of the plot, is then verified by the fact that in the figure 6, in the histogram, the residuals are more concentrated on the left side.

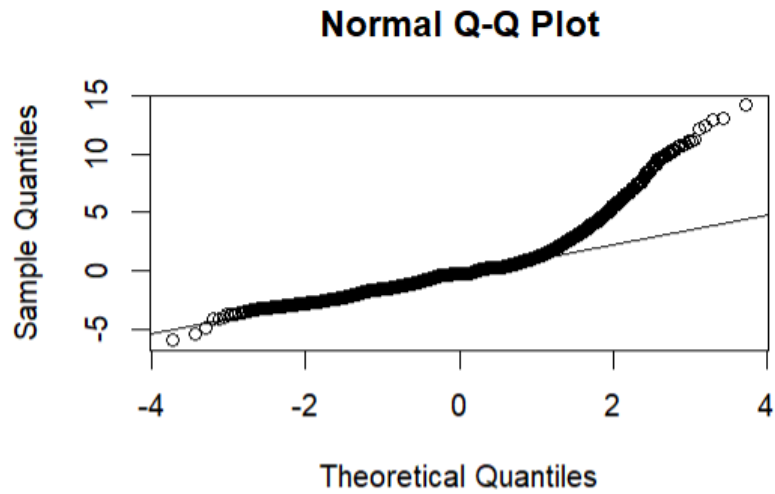


Figure 5: Q-Q plot

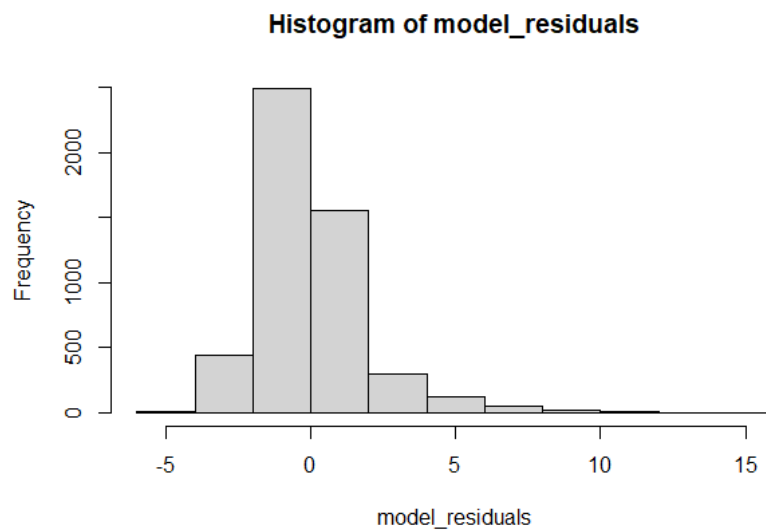


Figure 6: Histogram of Residuals

Being much like the Residuals vs Fitted plot, the Scale-Location plot simplifies the analysis of the variance of the dependent variable for all the data.

Below, the plot shows that the horizontal orientation of the red line suggests that the average size of the standardized residuals remains constant across the range of fitted values. Furthermore, the consistency of the spread around the red line shows that the magnitude variability does not vary much as a function of the fitted values.

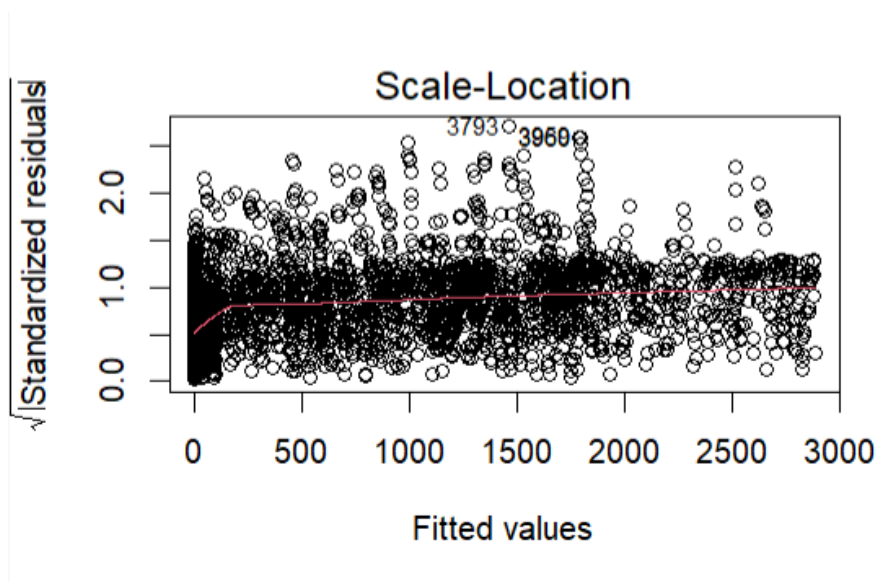


Figure 7: Scale-Location plot

<i>Performance measures</i>	
Multiple R	0,999997
R Square	0,999993
Adjusted R Square	0,999993
MAE	1,311842
MSE	3,972203
RMSE	1,993039

Figure 8: Performance Measures

Another way of reflecting on the model performance is to interpret multiple R-squared and adjusted R-square values. The R Squared and Adjusted R Squared

of the model values were respectively 0.999996693 and 0.99999338. These values indicate that the four independent values can explain approximately 99,999 % of the variation of the dependent variable. Also, MAE, MSE, RMSE values were calculated to measure the predictive capabilities of the model. Regarding MAE, the following value is 1.311842. This means that, on average, the model's predictions are deviated by 1.31 units from the actual values. MSE value is 3.972203 meaning that, on average, the model's predictions are off by about 3.97 squared units from the actual values. Finally, the RMSE is 1.993039 which indicates that the model's predictions are off by about 1.99 units from the actual values.

Overall, the model appears to function well, as seen by its extremely high r-squared value and moderately low error margin, being able to predict accurately the dependent variable, the flow time of productions. The regression model's parameters were examined, and it became apparent that the model complied with the fundamental assumptions of linear regression analysis. The model specifically showed a linear connection between the dependent variable and the independent variables, and it was discovered that each independent variable was found to be independent from the others. Furthermore, the errors had a normal distribution, and their variance was constant across all levels of the independent variables.

6. Conclusion

In conclusion, this thesis explored the problem of assigning due dates to jobs in a manufacturing company. The research questions centered on using a linear regression model to estimate the flow time of productions and determining the respective independent and dependent variables. In a whole, the most important results indicate that not only the independent variables from the model are able to significantly explain the response variable, but also every performance measure used to assess the model showed that it had low error margin.

The goal of this research was to create a linear regression model for estimating the flow time of production. The use of linear regression was chosen due to its practical importance and simplicity of implementation when compared to non-linear ones. The model was created and put through several tests using simulated data, and the numerical outcomes showed that it provided a high degree of accuracy. The lack of actual data, however, is the study's limitation. When attempting to apply the model to real-world data from a manufacturer, Grohe, it was found that the available data were inadequate to conceive a significant linear regression model. It's essential to be aware of this limitation because it emphasizes how crucial it is to collect accurate data that is adequate for model development and validation.

As a recommendation for manufacturing companies, data quality control methods should be implemented to guarantee that the information gathered is precise and reliable, including regular data cleaning and validation checks. Future research should make use of real data from manufacturing organizations to overcome this limitation to enhance the estimation of the flow time of manufacturing companies. To increase the precision of projections, adding potential variables that might affect the flow time of productions should be further investigated.

Bibliography

- Alpay, S., & Yüzügüllü, N. (2009). Dynamic job shop scheduling for missed due date performance. *International Journal of Production Research*, 47(15), 4047–4062. <https://doi.org/10.1080/00207540802043980>
- Amponsah-Kwatiah, K., & Asiamah, M. (2020). Working capital management and profitability of listed manufacturing firms in Ghana. *International Journal of Productivity and Performance Management*. <https://doi.org/10.1108/IJPPM-02-2020-0043>
- Chang, F. C. R. (1994). A study of factors affecting due-date predictability in a simulated dynamic job shop. *Journal of Manufacturing Systems*, 13(6), 393–400. [https://doi.org/10.1016/0278-6125\(95\)90063-Q](https://doi.org/10.1016/0278-6125(95)90063-Q)
- Chatterjee, S., & Hadi, A. S. (2012). *Regression Analysis By Example Fifth Edition*.
- Chen, Z.-L. (1996). Scheduling and common due date assignment with earliness/tardiness penalties and batch delivery costs. *European Journal of Operational Research* (Vol. 93) [https://doi.org/10.1016/0377-2217\(95\)00133-6](https://doi.org/10.1016/0377-2217(95)00133-6).
- Chen, Z. L., & Hall, N. G. (2008). Maximum profit scheduling. *Manufacturing and Service Operations Management*, 10(1), 84–107. <https://doi.org/10.1287/msom.1060.0144>
- Cheng, T. C. E. (1988). Optimal common due-date with limited completion time deviation. *Computers and Operations Research*, 15(2), 91–96. [https://doi.org/10.1016/0305-0548\(88\)90001-9](https://doi.org/10.1016/0305-0548(88)90001-9)
- Cheng, T. C. E., & Jiang, J. (1998). Job shop scheduling for missed due-date performance. *Computers and Industrial Engineering*, 34(2–4), 297–307. [https://doi.org/10.1016/s0360-8352\(97\)00317-3](https://doi.org/10.1016/s0360-8352(97)00317-3)
- Ciurana, J., Quintana, G., & Garcia-Romeu, M. L. (2008). Estimating the cost of vertical high-speed machining centres, a comparison between multiple regression analysis and the neural networks approach. *International Journal*

- of *Production Economics*, 115(1), 171–178.
<https://doi.org/10.1016/j.ijpe.2008.05.009>
- Demir, H. I., & Erden, C. (2020). Dynamic integrated process planning, scheduling and due-date assignment using ant colony optimization. *Computers and Industrial Engineering*, 149. <https://doi.org/10.1016/j.cie.2020.106799>
- Edh Mirzaei, N., Hilletoft, P., & Pal, R. (2021). Challenges to competitive manufacturing in high-cost environments: checklist and insights from Swedish manufacturing firms. *Operations Management Research*, 14(3–4), 272–292. <https://doi.org/10.1007/s12063-021-00193-0>
- Erden, C., Demir, H. I., & Kökçam, A. H. (2019). Solving Integrated Process Planning, Dynamic Scheduling, and Due Date Assignment Using Metaheuristic Algorithms. *Mathematical Problems in Engineering*, 2019. <https://doi.org/10.1155/2019/1572614>
- Evans, D. S. . (1987). The Relationship Between Firm Growth , Size , and Age : Estimates for 100 Manufacturing Industries Author (s): David S . Evans Source : The Journal of Industrial Economics , Vol . 35 , No . 4 , The Empirical Renaissance in Industrial Economics (Jun ., . *The Journal of Industrial Economics*, 35(4), 567–581. <http://www.jstor.org/stable/2098588>
- Garza-Reyes, J. A., Kumar, V., Chaikittisilp, S., & Tan, K. H. (2018). The effect of lean methods and tools on the environmental performance of manufacturing organisations. *International Journal of Production Economics*, 200, 170–180. <https://doi.org/10.1016/j.ijpe.2018.03.030>
- Gogtay, N. J., & Thatte, U. M. (2017). Principles of correlation analysis. *Journal of Association of Physicians of India*, 65(MARCH), 78–81.
- Gordon, V., Proth, J. M., & Chu, C. (2002). A survey of the state-of-the-art of common due date assignment and scheduling research. *European Journal of Operational Research*, 139(1), 1–25. [https://doi.org/10.1016/S0377-2217\(01\)00181-3](https://doi.org/10.1016/S0377-2217(01)00181-3)

- Gordon, V. S., Proth, J. M., & Chu, C. (2002). Due date assignment and scheduling: SLK, TWK and other due date assignment models. *Production Planning and Control*, 13(2), 117–132.
<https://doi.org/10.1080/09537280110069621>
- Gordon, V. S., & Strusevich, V. A. (2009). Single machine scheduling and due date assignment with positionally dependent processing times. *European Journal of Operational Research*, 198(1), 57–62.
<https://doi.org/10.1016/j.ejor.2008.07.044>
- Heizer, J., Render, B., & Munson, C. (2017). *Operations Management: Sustainability and Supply Chain Management* (Twelfth Edition). Pearson.
- Inman, R. A., Sale, R. S., Green, K. W., & Whitten, D. (2011). Agile manufacturing: Relation to JIT, operational performance and firm performance. *Journal of Operations Management*, 29(4), 343–355.
<https://doi.org/10.1016/j.jom.2010.06.001>
- Jacobs, R. F., & Chase, R. (2018). *Operations and Supply Chain Management* (Fifteenth Edition). McGraw-Hill.
- Janiak, A., Janiak, W. A., Krysiak, T., & Kwiatkowski, T. (2015). A survey on scheduling problems with due windows. *European Journal of Operational Research*, 242(2), 347–357. <https://doi.org/10.1016/j.ejor.2014.09.043>
- Karuppusamy, N. S., Pal Pandian, P., Lee, H. S., & Kang, B. Y. (2015). Tool wear and tool life estimation based on linear regression learning. *2015 IEEE International Conference on Mechatronics and Automation, ICMA 2015*, 17–21.
<https://doi.org/10.1109/ICMA.2015.7237449>
- Mokhtari, G., & Kianfar, F. (2007). Simultaneous Due date assignment and lot sizing with uncertain flow times. *International Journal of Engineering, Transactions A: Basics*, 20(3), 263–280.
- Seidmann, A., & Smith, M. L. (1981). Due date assignment for production systems. *Management Science*, 27(5), 571–581.
<https://doi.org/10.1287/mnsc.27.5.571>

- Shabtay, D., & Steiner, G. (2007). Optimal due date assignment and resource allocation to minimize the weighted number of tardy jobs on a single machine. *Manufacturing and Service Operations Management*, 9(3), 332–350. <https://doi.org/10.1287/msom.1060.0139>
- Song, D. P., Hicks, C., & Earl, C. F. (2002). Product due date assignment for complex assemblies. *Int. J. Production Economics* (Vol. 76).
- Teymourifar, A., & Ozturk, G. (2018). New dispatching rules and due date assignment models for dynamic job shop scheduling problems New dispatching rules and due date assignment models for dynamic 303. *Int. J. Manufacturing Research* (Vol. 13, Issue 4).
- Zhao, R., Liu, Y., Zhang, N., & Huang, T. (2017). An optimization model for green supply chain management by using a big data analytic approach. *Journal of Cleaner Production*, 142, 1085–1097. <https://doi.org/10.1016/j.jclepro.2016.03.006>

Appendix

Source	Purpose	Method	Findings	Limitation	Possible Future Research
Edh Mirzaei et al., 2021	Development of a compounded perspective of issues affecting firms' internal and external environments	Qualitative case study	Seven key categories of challenges identified	The outcomes are particular to a specific high-cost area	-
Jacobs & Chase, 2018	This book is about concepts related to operations and supply chain management	-	-	-	-
Cheng, 1988	Find the most efficient assignment sequence to minimize a penalty function for missed due dates	Common due-date assignment method	The due date is divided in 3 lemmas; numerical example is presented	-	TTW and SLK should be investigated
Mokhtari & Kianfar, 2007	The literature leaves out the lead time and due date-related expenses	Real-coded genetic algorithm.	resulting numerical findings	-	Demand can be considered random

Z.L. Chen, 1996	Common due date for all jobs must be established.	Polynomial dynamic programming algorithm.	Presentation of a objective function measured by the sum of earliness-tardiness penalties	The tardiness penalty weight is greater than or equal to the earliness penalty weight	An answer is needed for the fact that Earliness penalty weight could be higher than Tardiness penalty weight
Seidmann & Smith, 1981	Minimizing the expected aggregate cost per job subject	Algorithm	-	The optimal lead time is a unique minimum point of strictly convex function	-
V. S. Gordon & Strusevich, 2009	Problems related to single machine scheduling and due date assignment	CON and SLK methods polynomial-time dynamic programming algorithms of CON and SLK methods	Problem was solved with both algorithms	-	Other rules can be added to the dynamic programming framework
Z. L. Chen & Hall, 2008	Fill the gap left by scheduling literature, taking into account revenue and net profit maximization goals	Computational Algorithm	Effective scheduling in increases revenue as well as cost management	-	Study algorithms with na increased computation time
Garza-Reyes et al., 2018	The effects of lean methodology on environmental performance	Correlation analysis	Total Productive Maintenance and JIT have the strongest impact on environmental	Some constraints restricted the depth of the study	More sensible plans are required to balance environmental performance

			performance		with lean methodology
Inman et al., 2011	Investigation of the relationship between manufacturing agility and operational and firm performance.	A structural model tested; survey made on managers from American manufacturing companies	JIT-purchasing and agile have a relationship	Non-response bias	Relations of JIT to manufacturing agility and agility to performance.
Janiak et al., 2015	A survey of scheduling problems with due windows	Review of the literature	-	-	Review of research about scheduling with multiple due windows
Song et al., 2002	Investigation of the problem of assigning product due dates for complex multistage assemblies	Method for assigning product due dates	The method is effective	-	-
V. S. Gordon et al., 2002	Review for the due date determination models depending on the processing times	Survey of the literature	-	-	-
Erden et al., 2019	Study of the IPPSDDA	Metaheuristic algorithms	The algorithms showed to be competitive and generated solutions to	-	Dynamic events should be studied machine problems, job cancellation and arrival of

			the problem		urgent jobs
Shabtay & Steiner, 2007	Study a SC scheduling problem in a single machine system environment to schedule realistic delivery dates	polynomial-time optimization algorithm to minimize a cost function	-	-	Problems with different objective functions, different DDA methods
V. Gordon et al., 2002	Survey of the literature, involving the DDA models related to single and parallel machines	Survey of the literature	-	-	-
Cheng & Jiang, 1998	In the literature, missed due date performance has not been explored. Investigation is made on dispatching rules and DDA methods related to the gap.	Dispatching rule; 2 due date setting models developed	Dynamic proposed models are better than the stochastic ones	-	Study models related to flow time estimation
Alpay & Yüzügüllü, 2009	MAL and MSL considered as performance criteria to improve missed due dates	DDA model	The model was successful and improved the missed due date performance	-	Flowtime estimation.

Demir & Erden, 2020	Minimizing earliness, tardiness and due dates of jobs by finding the best combination of the dispatching rules	Eight different size shop floors were developed to compare GA and ACO algorithms	ACO performed GA algorithm	-	Makespan could be incorporated in a programming model
Teymourifar & Ozturk, 2018	To avoid earliness and lateness, accurate methods for DDA and more suitable rules for DJSSPs should be found	DDA models	The approach "estimation of the total waiting time in queues and adding it to the release time and processing time of each job." was the one with best performance	-	Introducing stochastic breakdown of machines and repair times with buffer conditions into new dispatching rules and models
Chang, 1994	It is found very hard to predict accurately the factors that affect jobs that enter in the production site in a random manner.	Method to identify these factors	Production managers should create DDA rules that have dependability on the characteristics of their production systems	-	-
Zhao et al., 2017	Solve a multi-objective optimization model related to GSCM	Three scenarios were proposed to solve the model	The optimization applicability is expected to determine the strategy that fits best the reduction of carbon	-	More big data analytical models should be used

			emissions		
Chatterjee & Hadi, 2012	The book addresses regression analysis topics such as simple and multiple linear regression	-	-	-	-
Evans, 1987	Relationship between three aspects of a manufacturing company growth: dissolution, variability of growth, three key characteristics of the company	Estimation of these relationships	Firm age is an important determinant of firm dynamics and the growth rate of a firm decreases at a diminishing rate with its size.	-	Use of datasets with experience information on companies
Ciurana et al., 2008	Understand which machine is more suitable for the necessities of a manufacturing company	Development of 2 cost models for vertical high-speed machining (HSM)	ANN gave an accurate estimation results than the MRA	Data unavailability on some key factors	-
Karuppusamy et al., 2015	Study on how some tools, just like the sharpened edge ones, may lose their sharpness	System to recognize tool wear predict the value of feed rate and depth of cut.	Method showed good performance	-	Study more manufacturing firms in case data is available

	during their use on production jobs				
Amponsah-Kwatiah & Asiamah, 2020	Examination of the relationship between working capital management on profitability in Ghana manufacturing listed firms	Correlation analysis	ROA and ROE have relation with inventory management , current ratio and others.	-	Exploration of the relation between working capital management and internal control systems
Gogtay & Thatte, 2017	This article presents the concepts related to correlation analysis	-	-	-	-

Table 5: Literature Review