ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis – Urban, Port and Transport Economics

Congestion in the Reversed Order Picking System of Royal FloraHolland

Kaj Overdevest

481320

Rotterdam, 27-11-2019

Supervisors:
Prof. dr. ir. Rommert Dekker, *Erasmus University Rotterdam (ESE)*
Frank Ebskamp, *Royal FloraHolland*

Second assessor:
Niels van Saase, *Erasmus University Rotterdam (centre for UPT)*

ACKNOWLEDGEMENTS

In front of you lies a thesis conducted for the Master's programme Urban, Port and Transport Economics that has received great effort and required a lot of hard work. I could never have achieved this without the help of many people, whom I would like to thank in this preface.

First of all, I would like to show my appreciation to my university supervisor, Rommert Dekker. His widespread knowledge in warehouse operations has provided me with great insights and helped me in finding the best direction for my thesis research.

Furthermore, without Royal FloraHolland this thesis would never have existed. Specifically, I would like to thank Frank Ebskamp and Sandra van den Berg, my company supervisor and his colleague. Their expertise and company-specific knowledge have been of great importance and gave me the possibility to conduct the right research for this thesis.

In addition, my girlfriend and parents have been a listening ear whenever I wanted to talk about my research and the problems I encountered, for which I am very grateful.

I wish the reader a lot of fun reading this thesis and wisdom with the knowledge that it provides.

ABSTRACT

In *picker-to-parts* order picking systems interactions between workers occur as *pick face blockings*, indicating that two workers try to access the same location at the same moment in time. Existing literature on warehouse congestion mostly considers theoretical modelling, making it difficult for warehouse managers to assess their operations' performance. Therefore, this thesis bridges the gap between theory and practice by creating real-life insights in warehouse congestion through a data-driven approach.

The case study on a reversed order picking system shows that data mining is required in order to retrieve valuable information on warehouse congestion. This data mining enables the identification of *pick face blockings* and quantification of the time that is lost due to congestion accordingly. Moreover, regression analysis shows that congestion is caused by the number of stops made by workers in the system, total time spent on transferring products and the number of locations that has been accessed.

MANAGEMENT SUMMARY

Problem Statement

This thesis focused on quantifying the waste due to congestion in reversed order picking systems and identifying its main causes. Specifically, a case study was conducted on Royal FloraHolland's distribution process of flowers in their warehouse in Naaldwijk, the Netherlands. In this *picker-to-parts* warehouse, workers drive along the aisles in order to transfer the products they are carrying to their designated storage locations. Royal FloraHolland has reached its warehouse limits in terms of the number of workers they can deploy, taking safety issues into account. Consequently, the warehouse is overcrowded and congestion plays a vital role each and every day. This congestion translates into *pick face blockings*, which are situations in which multiple workers try to access the same storage location at the same point in time, causing waiting times. Even though managers at Royal FloraHolland are familiar with this problem, quantifying the time they lose has proven to be a challenge that has not yet been solved. Also, since existing literature on warehouse congestion mostly concerns theoretical modelling, this thesis aimed at creating valuable insights from practice.

To overcome this problem, the following main research question central to this thesis was introduced:

Where does congestion occur in the reversed order picking system of Royal FloraHolland and what are its most important causes?

This research question was divided into four sub-objectives: visualization, quantification, identification and regression.

Results

The *visualization* of the stops made by the workers showed that the warehouse is characterized by an unequal distribution of activity. Extensive data mining-practices exposed the patterns of *pick face blockings* throughout the warehouse, confirming this unequal distribution. A heatmap tool was developed that gives Royal FloraHolland the opportunity to visualize their process and locate the occurrence of congestion at a glance.

Quantification of the congestion took place by setting up theoretical travel times and comparing these with actual travel times. Accordingly, it turned out that workers that had experienced pick face blocking lost nearly 20 seconds, on average. Analysis of the complete distribution process of August 26 showed that Royal FloraHolland had lost 30 manhours due to congestion on that day. With this quantification method in place, Royal FloraHolland can actively monitor the amount of waste they incur.

The *identification* of different explanatory variables for the number of pick face blockings in the warehouse was done through literature study and quantitative analysis of the data made available by the data mining process. *Regression* of all identified variables resulted in a model that was able to explain 94% of the variability in the number of pick face blockings. This model included the number of stops made by the workers, total time spent on transferring products and the number of locations that were utilized. With these three parameters, Royal FloraHolland was given a direction for future research that has to be conducted in order to reduce congestion. Moreover, the regression model can be used to predict an expected number of pick face blockings in the process for comparison with reality.

TABLE OF CONTENTS

LIST OF ACRONYMS

RFH	Royal FloraHolland
OLS	Ordinary Least Squares
ZCM	Zero Conditional Mean

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

This chapter starts with an introduction of the problem and research question central to this Master's thesis. The second section gives detailed information on the case study at hand. Section 1.3 states the research sub-objectives for this research and Section 1.4 outlines this thesis' structure.

1.1 Problem Statement

In warehouses, products arrive in large clustered quantities, remain stored until ordered and leave packed in smaller quantities. But warehouses are more than just a place to stock products as they are an integral part of large supply chains involving numerous parties. The performance of such supply chains heavily depends on warehouses and their related activities. Inventory management, reducing lead times and performing value added services within these warehouses contribute to the performance and resilience of supply chains as a whole.

Order picking is one of the major processes in warehouses, accounting for up to 55% of total operating costs (De Koster, Le-Duc, & Roodbergen, 2007) and can be described as "*the activity by which a small number of goods is extracted from a warehousing system to satisfy a number of independent customer orders*" (Goetschalckx & Ashayeri, 1989). Each individual order concerns a very small percentage of the total number of goods stored in a specific warehouse. Order picking is known as a very labour intensive operation and as a capital intensive operation for automated systems. Order picking includes various activities, such as clustering and scheduling customer orders, assigning stock locations to order lines, releasing orders from the system, picking the products from the different storage locations and disposal of the picked products (De Koster et al., 2007). In standard order picking operations, one customer order involves multiple order lines with the related stock keeping units, quantities and locations in the warehouse.

The picking operation of these order lines can be organized in various ways. To start with, order picking can roughly be split into two systems, the first being mainly operated by manual labour and the second type of system being largely automated. Most warehouses that deploy manual labour are known as *picker-to parts* systems. In these systems, the order picker walks or drives along the aisles in the warehouse to retrieve the requested products from the storage locations (De Koster et al., 2007). Picker-to-parts systems can be divided into two subsystems:

low-level and *high-level* picking. Low-level picker-to-parts systems are designed in a way that allows pickers to reach all storage locations from the floor. In high-level picker-to-parts systems, pickers use a lifting order-pick truck or crane to perform their picking activities. Systems with cranes are semi-automated as they automatically stop in front of the designated warehouse location and wait for the picker to accomplish their job (De Koster et al., 2007).

As order picking is a tedious task in classic picker-to-parts systems and most of the distribution centers still belong to the labour-intensive industry, process optimization is crucial (Hsieh & Tsai, 2006). A major issue affecting the efficiency of workers in order picking systems is waste as a result of congestion. The following expression from Huber (2014) can be used to clarify congestion: "*In manual order picking systems we speak of congestion whenever the activities of an order picker are interrupted by another order picker while all requirements for a regular picking process are met*". Because most order picking systems are utilized by multiple workers at the same time, interactions are unavoidable. The blocking situations that occur as a result of these interactions are a potential threat to service levels and put pressure on operating costs. The most important factors that influence congestion are the number of workers and the width of aisles (Huber, 2014)**.** However, decreasing the number of workers lowers total throughput in terms of order lines and increasing the width of aisles to reduce congestion comes with huge costs of surface. Therefore, managers should strive to optimize their warehouse by making a trade-off between the costs of congestion and the costs of surface.

Order picking systems are heavily discussed in existing literature, mostly focusing on warehouse layout and storage and routing policies. Some attention has been paid to congestion through simulation studies and analytical models, however, scientific literature has hardly integrated real-life data. Integrating simulation or analytical models is a very time-consuming process and therefore often skipped by warehousing companies. Consequently, congestion consideration is often omitted from warehouse optimization by managers even though it has major effects on productivity. The question that remains is: why is congestion hardly quantified in existing literature? Is this practice too complicated or are companies simply overlooking the possibilities that come with congestion quantification? Therefore, this thesis focuses on exploring and quantifying congestion in order picking systems through a data-driven case study.

Data from the reversed order picking system of Royal FloraHolland is used to bridge the gap between theory and practice. Moreover, the uniqueness of the system provides valuable insights for further scientific research. This thesis will carry out congestion quantification in practice, which has rarely been done before. Next to the quantification of congestion, the research will incorporate analysis of the root causes of congestion. This thesis focuses on answering the following main research problem:

Where does congestion occur in the reversed order picking system of Royal FloraHolland and what are its most important causes?

1.2 Royal FloraHolland Case

The cooperation Royal FloraHolland (RFH) has been a worldwide marketplace for flowers and plants for over a hundred years. The basics of economics, supply and demand from growers and buyers, are at work in the auction halls of RFH every day. Optimal pricing is able to take place as RFH strives to keep transaction costs as low as possible for its clients. Doing so, product flows have to be taken care of on both the supply and demand side.

Flowers are brought to the auction by their growers on transportation units called *stapelwagens*, which will be referred to as roll containers from now on. After the flowers are sold in the auction process, they have to be distributed to their buyers. Buyers are located in-house and since RFH's facility is of a large scale, products are collected in a buffer zone before they are delivered.

Prior to this buffer zone, the distribution process takes place. Up to 200 workers make sure that after the flowers passed the auction process, flowers on almost 7000 roll containers are handled and distributed to so-called client-specific roll containers on a daily basis. These client-specific roll containers act as a temporary storage location for the products. After a client-specific roll container is filled up with products, it is displaced to the buffer zone so that it can be picked up for final delivery. All workers travel through the warehouse on electric trucks as shown in Figure 1.1.

Figure 1.1 – Two aisles with electric trucks

In the distribution process, a roll container that has left the auction process can contain products for multiple clients. This implies that a worker has to make several stops to distribute the products to the client-specific roll containers. Hence, the distribution process can be considered as a reversed order picking system. Instead of picking products from storage locations in a warehouse, workers at RFH are responsible for distributing products from the roll container to their client-specific storage locations.

Roll containers pass through the auction process one by one and are collected in a buffer zone afterwards. Each worker picks up a roll container from this post-auction buffer zone and accomplishes his route with a number of stops at client-specific storage locations. These locations are designed in a way that is similar to warehouse aisles for order picking. On both sides of the aisles, workers can stop to transfer products to the client-specific roll containers.

Within this process RFH is bothered with congestion, which causes waiting times and most likely results in excessive working hours. RFH wants to know when congestion occurs to better understand the distribution process. Managers at RFH are faced with low worker productivity rates which they cannot explain, probably caused by congestion. Moreover, managers have no clue of how many manhours they lose on congestion. All of the managers involved in the distribution process unquestionably agree on the fact that congestion exists, even though not a single one of them is able to quantify the costs that are incurred accordingly. The major issue for this thesis is to visualize and estimate the congestion for cost analysis and identify the causes of congestion as input for management decision-making and future academic research.

1.3 Research Sub-Objectives

Based on the RFH-case, four specific sub-objectives have been identified to produce a proper answer to the main research question. Moreover, each sub-objective has its own research question and corresponding chapter within this thesis.

1.3.1 Visualization

Interpretation of the different data sets to visualize volatility of the process. This sub-objective aims at creating a clear understanding of the case study at hand. Visualization of the different datasets depicts the situation at RFH and provides a decent fundament for this thesis research. Chapter 5 deals with the following question: What does the data from RFH look like and how can it be used to visualize the process as a start of congestion analysis?

1.3.2 Quantification

Setting parameters for theoretical travel times to quantify congestion. As the datasets from RFH contains meaningful timestamps, Chapter 6 aims at modifying the data to set up theoretical travel times. These theoretical travel times will form the basis for congestion quantification. This chapter answers the following question: How long does it take the workers to drive to locations, on average, on a day without congestion and how much time is wasted accordingly?

1.3.3 Identification

Chapter 7 is a start in congestion analysis and explores the possible causes of congestion in the RFH case study. What could be the most important causes of congestion in RFH’s reversed order picking study? Through literature review, data analysis and interviews with managers at RFH, this question will be answered.

1.3.4 Regression

After the most important influencers of congestion are identified, their statistical relevance will be tested through regression analysis. This regression analysis aims at confirming the true importance of several congestion root causes. Therefore, Chapter 8 deals with the following research question: What is the best regression model to explain congestion in RFH’s reversed order picking system?

1.4 Thesis Outline

This thesis continues with information on the methods that were used and a thorough data description in Chapter 2. This chapter also includes detailed information on the case study at hand. In Chapter 3, all relevant literature on order picking congestion is reviewed. Moreover, this chapter identifies the relevant contribution of this thesis to the existing literature. Chapter 4 describes the methods that were used to answer the main research question. The following four chapters are named after the sub-objectives formulated in the previous section: Chapter 5 visualizes congestion in the reversed order picking system, Chapter 6 aims at an objective quantification of congestion, the most important influencing factors of congestion are identified in Chapter 7 and then used to explain and predict congestion in Chapter 8. Based on its predecessors, Chapter 9 states the main research findings and provides recommendations regarding the case study. Chapter 10 discusses this thesis' limitations and points out possible directions for further scientific literature. Finally, Chapter 11 underlines the importance of this thesis' research for RFH by emphasizing on the steps that managers have taken to implement the newly developed methods.

CHAPTER 2 DETAILED PROBLEM DESCRIPTION

This chapter provides a thorough explanation of the case study at hand for this thesis. Section 2.1 describes the reversed order picking system in detail through different sub-sections on warehouse layout, routing policy, storage location strategy and other relating activities. All relevant datasets that are used for this thesis are outlined in section 2.2. Based on the detailed case description and the available datasets, section 2.3 concludes the most important takeaways and implications for this research.

2.1 Case Study

The case study used for this thesis is based on RFH's warehouse in Naaldwijk, the Netherlands. This specific location of RFH has no room for further expansion, resulting in more congestion compared to other facilities. The distribution process of flowers that are sold at the auction starts every weekday at six o'clock in the morning and lasts until all products are delivered to their new owners, which is usually five hours after the start of the process. Total throughput fluctuates throughout the year and more importantly, per weekday. Consequently, the number of workers in the warehouse changes and congestion plays a vital role on the busiest days of the year. For safety reasons, managers have set the absolute limit at 195 workers driving through the warehouse. But with limited space available and no option to expand the warehouse on short term, congestion is unavoidable. A lack of research concerning this congestion created a blind spot for RFH's managers. Therefore, investigation proving that congestion costs money should contribute to convincing the executive board of the necessary warehouse improvements.

2.1.1 Warehouse Layout

Figure 2.1 shows a simplified top view of RFH's warehouse in Naaldwijk. The blue marked areas indicate the client-specific storage locations which are spread over 24 aisles. Workers in the distribution process pick up roll containers from one of the two buffer locations at the East- and Westside of the warehouse and then start their trip in the direction of the arrows. Final delivery buffer zones are located at the end of the aisles and the warehouse includes designated areas for special deliveries ('EEN- and RESTKOOP transactions') which will be discussed later on in this chapter.

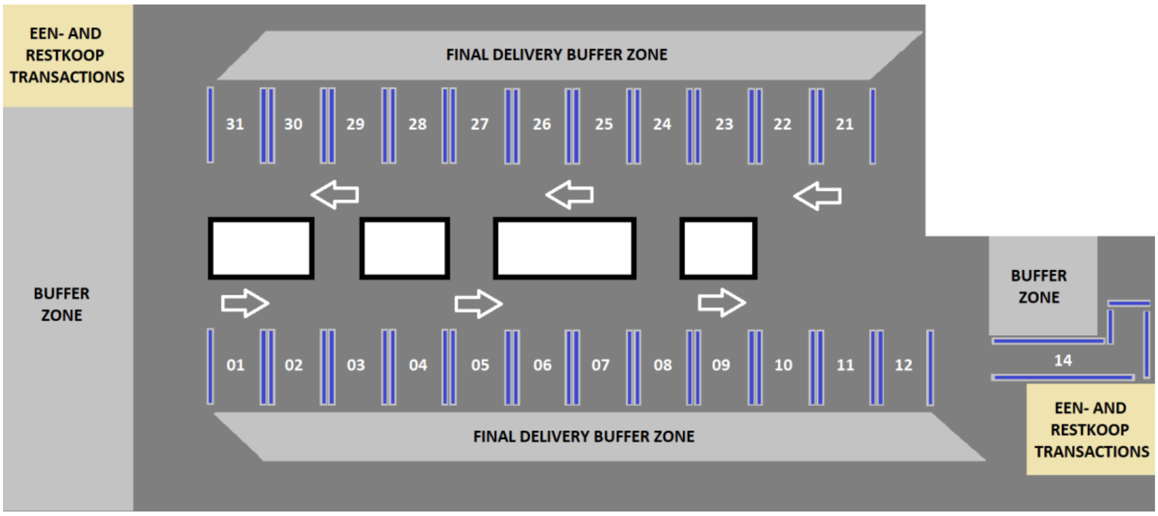

Figure 2.1 – Warehouse layout

2.1.2 Routing Policy

The distribution process starts at the pickup point from one of the two post-auction buffer zones. Depending on the location of the buffer zone, a worker starts his journey at the East- or Westside of the warehouse when he picks up a roll container. When starting at the Westside of the warehouse, a worker comes across aisle 1 first until he makes a turn at the end of the cross aisle (at aisle 12 and 21). Starting from the Eastside, a worker could enter aisle 21 first until he makes a turn at the end of the cross aisle (at aisle 31 and 1). Workers are only allowed to drive in anti-clockwise direction to prevent traffic accidents in the warehouse. RFH works with a voice picking system that tells the worker which aisle he has to enter from the cross aisle to transfer products at a client location. Accordingly, it can be concluded that RFH uses the return method for order picking (De Koster et al., 2007). Using this method means that each worker visits an aisle, if and only if, a stop has to be made in it whilst following a fixed route through the warehouse. However, the route is not completely fixed since workers have three shortcut options to traverse from one side of the warehouse to the other. Besides the crossings at both ends of the warehouse, workers can traverse at aisle 3, 5 and 8. Accordingly, workers can decrease their travel distance if they do not have to make stops in certain aisles.

The aisle numbers 1 up to and including 12 and 21 up to and including 31 can be reached from the regular cross aisle. Some client locations form an exemption from the regular process, for example, aisle 14 is located at the very end of the distribution process. Moreover, products that are destined for the EEN- and RESTKOOP areas do not enter the regular distribution process as these are special deliveries.

2.1.3 Storage Location Strategy

Since the number of buyers that is active in the auction halls fluctuates during a week, RFH needs to reconsider their allocation of client specific storage locations in the warehouse every day. Some very large export companies are present every day, however, smaller garden centers that maintain their own procurement may only buy flowers a couple of times per week. In total, there are around 450 buyers active in Naaldwijk. Unfortunately for RFH, they do not announce on which days they will be active. Therefore, a rough planning is made each and every day, called the *drukverdeling*. This planning is a form of forecasting and uses historical data of the same weekday from one and two weeks ago, as well as from one year ago. Next to that, the forecast of total number of roll containers to be auctioned and related transactions from the commercial department is used in the *drukverdeling*. Based on this input, the estimates of throughput are spread out in an attempt to come to a fair distribution of the workload over the 24 aisles in the warehouse.

2.1.4 Final Delivery

As final delivery is directly linked with the distribution process, it has to be accounted for in the *drukverdeling*. In the final delivery buffer zones at the end of all aisles, roll containers are lined up in separate rows for clients. First of all, the client locations in the buildings are divided into a North and South section. Clients in the North section are assigned with a storage location in the upper part of the distribution process, whilst clients in the Southern part of the final delivery process get a storage location in the lower part of the warehouse (Figure 2.2). Accordingly, storage locations are not interchangeable between the two sections of the warehouse.

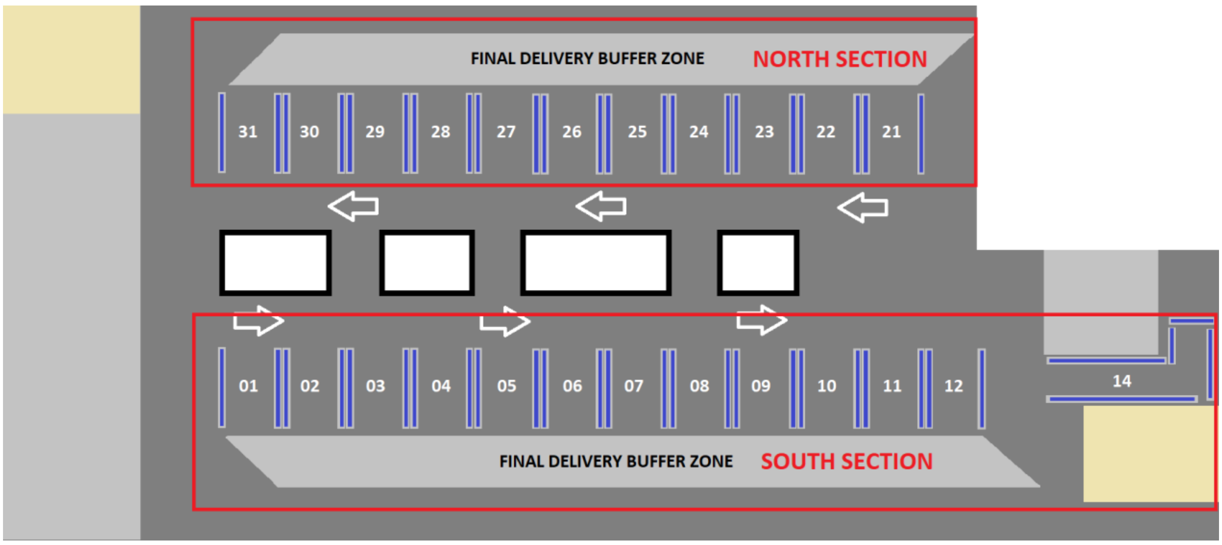

Figure 2.2 – Warehouse sections

Within both sections, clients are grouped based on the sub-buildings in which they are located. Moreover, final delivery knows several zones within each sub-building. These final delivery-zones are assigned to the buffer in a certain sequence. Final delivery is done with electric trucks and a single train may consist of up to fifteen roll containers (Figure 2.3).

Figure 2.3 – Electric truck for final delivery

Workers from the final delivery pick up multiple rows of roll containers at the buffer, whilst taking into account their drop-off sequence. This means that when a worker picks up five rows of roll containers, he picks up roll containers for five delivery zones within a sub-building or geographical area of the flower auction. When delivering, workers decouple the last roll containers from their train and continue to the next locations, which is why roll containers that have to be dropped off first are picked up from the final delivery buffer zone lastly.

To prevent traffic accidents, workers are only allowed to drive in Eastward direction when picking up rows of roll containers for final delivery. Consequently, the pick-up sequence from the buffer zone is fixed for clients from the same final delivery-zone. This has major implications for the *drukverdeling*. For example, clients located in final delivery-zone 1 are assigned with a drop-off number starting at 1, indicating whether roll containers will be dropped off first, second, etc. This implies that the client from delivery location 1.1 has to be picked up after the client from delivery location 1.2 is picked up. Consequently, if the latter is assigned a storage location in aisle 10, the client from delivery location 1.1 can only get a storage location in aisle 11 or 12. A particular trip is visualized in Figure 2.4, for which the worker will decouple roll containers at final delivery location 1.1 first, continues to location 1.2, until he is finished at the fifth location in this specific final delivery-zone.

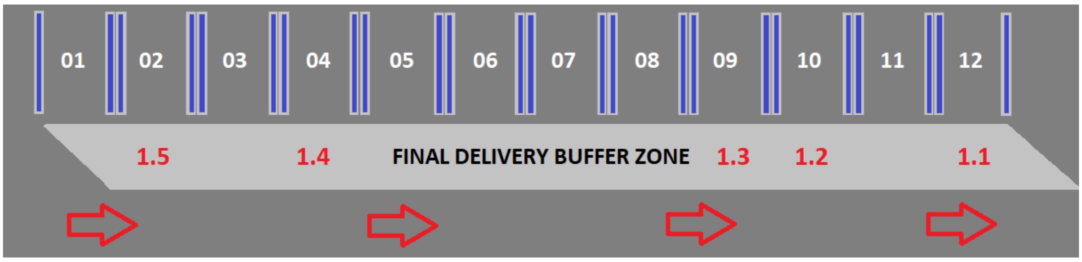

Figure 2.4 – Final delivery pick-up sequence

2.1.5 Order Picking Activities

As RFH's distribution process turns out to be a very special case of reversed order picking, evaluating and modelling it can be a tricky business. To caption all relevant aspects for further analysis, this section will discuss additional activities performed by the workers.

Checking

After a worker has picked up a roll container from one of the buffer zones, he is supposed to check it on product type, quantity and damage. The warehouse has two designated areas for this additional activity next to the buffer zones.

Replacing roll containers at client locations

At the start of the distribution process, all client locations are simply empty roll containers. After a number of visits by different workers, a client location may be fully loaded with flowers. When this happens, it has to be replaced by an empty roll container again. These empty roll containers are stored in the middle of the aisles, in order to allow the workers to drive around them and reach every client location. Each aisle has its own controller responsible for retrieving the fully loaded roll containers, moving them into the buffer zone for final delivery (*aflevergoot*) and providing the client location with a new empty roll container. However, the aisle controllers are not always capable of retrieving all of the fully loaded roll containers, which means that the workers have to spend time on replacing fully loaded roll containers with empty ones themselves.

Disposal of empty roll containers

After a worker has transferred all the trays from a roll container to their destined client locations, the empty roll container is all that is left over. As mentioned in the previous section, each aisle has designated space for empty roll containers. When a worker is left over with an empty roll container, he can drop it off in this designated area so that it can be used as client location at a later stage in the distribution process.

RESTKOOP transactions

Order lines that exceed a threshold in number of trays are classified as RESTKOOP, indicating that these trays do not have to be transferred at a client location. This type of transaction can remain on the roll container, which is then delivered at the end of the aisle that corresponds with the client location. The aisle controller immediately retrieves this roll container from the aisle and displaces it into the buffer zone to be picked up for final delivery. This saves workers a lot of time since trays do not have be transferred. On the other hand, as these roll containers are immediately retrieved from the aisles, final delivery may be done with roll containers that are not optimally loaded.

EENKOOP transactions

Next to RESTKOOP, the distribution process includes EENKOOP transactions. This means that all flowers on a roll container are bought by a single party. Usually, these transactions are delivered at the end of the aisle corresponding with the buyer's client location, similar to RESTKOOP transactions. However, some large buyers are known to engage in multiple EENKOOP transactions per day. To ease the pressure on the regular distribution process, these roll containers are delivered at two designated areas in the warehouse instead of at the end of the aisles. Note that RESTKOOP transactions may also be delivered in one of these designated areas to unburden the aisles.

2.2 Data

2.2.1 Commercial Dashboard

Flowers have to be sold in the auction halls before they can be distributed by RFH's workers in the warehouse. The commercial dashboard stores valuable information on the auctioning process in an Excel file. For each transaction, the commercial dashboard contains a row with important information for distribution of the related products. Each record in the commercial dashboard contains information on the delivery location, referring to a client specific roll container. The client specific roll containers are located in the warehouse aisles and have their own *distributienummer*, which is either a three- or a four-digit number. In the case of a three-digit number, the first digit refers to the aisle number (9 or lower) and the second and third digit refer to the location in the aisle. Regarding a four-digit number, the first two digits refer to the aisle number and the last two digits contain the location number.

The most relevant information for this research that is captured by the commercial dashboard is:

- *Veilgroep*; indicating the type of flower;
- *Distributienummer*; the buyer's location in the distribution process, a combination of aisle number and location in the aisle;
- *Kooptype*; separating the transaction types;
- *Stw nr. bij veilen;* unit number of roll container with flowers to be distributed;
- *#Fusten;* number of trays of the specific commercial transaction.

2.2.2 Trace Distributors

In practice, buyers may carry out multiple commercial transactions for a specific product category. Since the flowers are stored on the same roll container when being auctioned, these transactions are merged into a single logistical transaction. The trace distributors-file is a logged activity trail that contains solely logistical transactions, making it perfectly suitable for congestion analysis. The following variables are stored in the trace distributors-file:

- *Operator*; employee number that distributed the flowers;
- *LogTime*; timestamp created when worker received specific task;
- *Wagen*; unit number of roll container with flowers to be distributed;
- *Locatie*; location in the warehouse to which worker has to drive;
- *Partij*; number of product classifications on the roll container (based on dimensions and colors within the *Veilgroep*);
- *Aantal*; number of trays to transfer at specific location;
- *Restkoop*; indicating whether specific transaction is a RESTKOOP, as described in the previous section;
- *TaskStatus*; optional voice message created by worker.

Besides the regular information about storage locations and quantities, the voice picking system of RFH includes a number of additional options that are stored in the TaskStatus variable. With respect to congestion in the distribution process of RFH, it is probably the most interesting to look into additional features that point towards delays. For example, workers can choose to tell the voice picking system that they want to skip an aisle or location. If an aisle or location is congested, the workers can simply continue to the next aisle in which they have to transfer products to avoid waiting time. Also, workers are able to inform the system when a storage location is full, meaning that they have to retrieve the client specific roll container from the aisle and replace it with an empty one. This is requires additional activity that could possibly lead to delays as workers will block locations in the aisle for a longer period of time.

2.2.3 Logfiles Voice Communication

As mentioned in the previous chapter, workers in the distribution process get their instructions through a voice picking system. Pick-by-voice or voice picking is a technology that makes use of audio and voice to guide the order picking process, with greater worker productivity being the ultimate goal (de Vries, de Koster, & Stam, 2016). All pickers are equipped with a small terminal on their belt and wear a headset which informs them about their next storage location at which they have to transfer flowers.

The voice logfiles are stored in text format and contain all data traffic between the headsets and the main terminal which assigns tasks to the workers. The text files may add up to 1.5 million rows for the distribution process of a single day and contain information that is hard to capture at first sight as can be seen in the following example:

2019-08-26 11:15:46,014|INFO |NaaldwijkDistributieAWS|3480823|10.16.93.65 |900853|200344345|900853437034706 4328|VocollectTerminalResponseWriterImpl|#1:<0, ,ResponseXdockLBL,20009008530,900853,20190826111545,%S%,,%S%,0,0,1,,,,0,0,,,%S%,, 0,0,,0,0,,%S%,,%S%,,%S%><CRLF>

However, after data manipulation with the programming language R, it was possible to retrieve a highly relevant datapoint for each observation: the arrival timestamp of a worker at a specific client location.

2.2.4 Worker Accumulation

A fourth data file is retrieved from RFH’s Warehouse Management System to take the number of workers that are active into account. The total workforce is divided into subgroups for the breaks to make sure that the distribution process never comes to a standstill. However, the workforce will still be reduced significantly, hence influencing the probability of congestion (Schrotenboer, Wruck, Roodbergen, Veenstra, & Dijkstra, 2017). The file counts all unique employee numbers that transfer products in timespans of fifteen minutes. Based on the assumption that it takes no longer than fifteen minutes to handle a single transaction, this should cover all active workers in the process.

2.3 Conclusion

The reversed order picking system of RFH is a very specific case, subject to many constraints. For example, clients cannot get a random storage location in the warehouse as the system is bound to a specified sequence for final delivery. Next to that, the distribution activities of the workers are sometimes interrupted because they have to provide a storage location with a new empty roll container. These issues complicate the comparison of the RFH case study with regular order picking systems and require strict specifications of the system for further research.

A satisfying amount of datasets is available for research on congestion in RFH's reversed order picking system. However, quantification of congestion will not be done in one go as the data requires manipulation first. Since not a single dataset yields hard evidence for the existence of congestion and relevant information is scattered throughout the different datasets, adaption is required in order to provide meaningful insights for analysis. These steps will be conducted in the subsequent chapters.

CHAPTER 3 LITERATURE REVIEW

Order picking systems have been widely discussed in academic literature. This chapter gives an overview of existing literature for complete comprehension of order picking theory. Section 3.1 defines the different order picking activities and the challenges that warehouse managers have to face. This section establishes the base knowledge for section 3.2, in which all types of congestion are outlined. Regarding the RFH case study, it is of key importance that all possible causes are identified for analyzation purposes. As discussed in Chapter 2, the RFH case study is a very unique example, indicating that perfectly matching literature is hard to find. Hence, the relevance of academic articles has to be taken into account and should be constantly reviewed during the research. Section 3.3 deals with congestion evaluation methods and their relevance for the RFH case study. Lastly, section 3.4 states the main takeaways from existing literature and their shortcomings with respect to the RFH case study.

3.1 Order Picking

Order picking systems are subject to many constraints, such as labour, machines and capital (Goetschalckx & Ashayeri, 1989). Under these constraints companies strive to optimize their service levels, which are influenced by lead times and order integrity and accuracy. In the end, each order picking system is part of a larger supply chain meaning that the faster an order can be picked, the sooner it can be delivered to the final consumer. Because of this direct link with consumer satisfaction, minimizing pick times is crucial to warehouse managers (Le-Duc, 2005). However, picking operations are put under enormous pressure due to the introduction of new operating programs such as *Just-In-Time delivery, cycle time reduction* and *quick response* (Tompkins, 2010). One major implication for order picking operations arising from these developments is that warehouses have to deal with smaller orders which require more frequent deliveries. Handling costs in warehouses generally increase when order sizes decrease (Bartholdi & Hackman, 2008). As a result, controllability of warehouse operations such as order picking has dramatically increased over time.

Order picking operations are often divided into five main activities: travel, search, pick, setup and other secondary activities. As shown in Figure 3.1, 50% of order picking time consist of travelling. Travel time between locations in the warehouse is a direct expense, but does not add value and hence can be categorized as waste (Bartholdi & Hackman, 2008). Waste always has to be reduced in order for companies to become more profitable and meet customer demands.

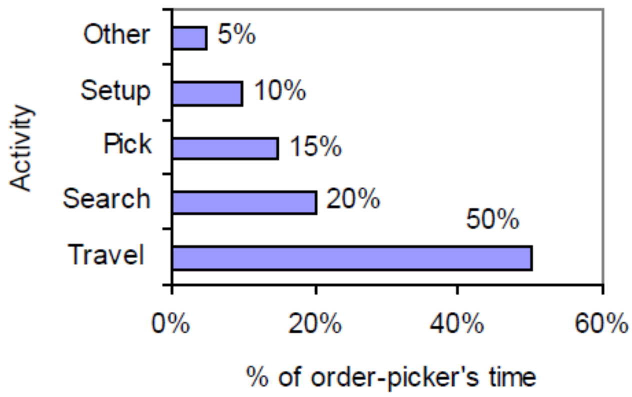

Figure 3.1 – Typical distribution of an order picker's time (Tompkins, 2010)

A realistic assumption in order picking systems is that travel time is an increasing function of total travel distance (Le-Duc, 2005). Reducing this travel distance to a bare minimum is known as the *Traveling Salesman Problem* and has much been studied in literature. Its goal is to find the cheapest way of visiting all cities or locations and returning to the starting point afterwards (Applegate, Bixby, Chvatal, & Cook, 2011).

Travel distance in order picking systems mainly depends on warehouse layout, aisle width and length, the storage assignment and routing methods (Le-Duc, 2005). For example, warehouses with multiple blocks (i.e. one or more middle cross aisles and rear aisles for traversing through aisles) have additional travel distance compared to one-block warehouses due to the distance that workers need to traverse when moving from one block to another. Also, total distance of a pick tour is affected by the locations that need to be visited. Storage assignment policies include several strategies to allocate products over the available locations in a warehouse. For instance, a random storage strategy locates products at random, meaning that each location along any aisle has the same probability of being visited. On the other hand, class-based storage strategies may result in longer or shorter average pick tour distances as certain groups of products are clustered in the warehouse.

Warehouses are limited to labour, machine and capital constraints. This implies that reducing total travel distances does not necessarily mean that an order picking system is operating at its full potential. Especially with order picking systems that require flexibility and scalability, these constraints can become a major issue. Warehouse operations that face significant variations in demand may want to increase the amount of labour that is put into the system to cope with the increased number of orders that has to be picked. However, when demand doubles in size and the workforce is doubled accordingly, this does not necessarily mean that all products will be retrieved from the system in time (Schrotenboer et al., 2017). This lower productivity is likely to stem from an increase in the number of workers in the system as more interactions between order pickers will occur. Interactions decrease the driving speed of workers and increase total travel time, already defined as waste by Bartholdi & Hackman (2008). Therefore, any type of congestion should be avoided as much as possible when optimizing order picking systems.

3.2 Congestion

Most scientific papers on order picking systems focus on single picker-to-parts systems which do not suffer from congestion. Fortunately, there are academic articles that consider multiple pickers. Huber (2014) gives an overview of existing models from other authors which include multiple pickers and related interdependencies that lead to congestion in order picking systems. With respect to congestion issues it is important to state that in narrow-aisle order picking systems, pickers can completely block an aisle. To the contrary, workers are able to pass each other in wide-aisle systems. The order picking system from the RFH case study is equipped with wide aisles, which is why relevant literature mainly on congestion in wide-aisle systems is discussed in this section.

It is important to note that not all waiting times in an order picking system stem from congestion. Orders that are coming in slowly and technical failures can also lead to waiting times. On the other hand, congestion can cause waiting times when order pickers are interfering with each other. Huber (2014) defines congestion in the following manner: "*In manual order picking systems we speak of congestion whenever the activities of an order picker are interrupted by another order picker while all requirements for a regular picking process are met*".

To understand where and when congestion might lead to waiting times in RFH's distribution process, it is important to assess the blocking situations that have been defined in existing literature. In his article, Huber (2014) distinguishes the following situations:

1. *Pick Face Blocking* – when two or more pickers want to access the same pick face at the same time. (Client-specific roll containers in the RFH-case study).
2. *In-the-Aisle Blocking* – occurs in narrow-aisle systems in which pickers cannot pass each other in the picking aisles.
3. *In-the-Aisle Interferences* – in wide-aisle systems, pickers can pass each other but still interfere as they will have to adapt their travelling speed or make sideward movements.

4. *Total Aisle Blocking* – if a certain picker blocks a whole picking aisle preventing other pickers from utilizing it.
5. *Cross Aisle Blocking* – similar to in-the-aisle blocking situations, cross aisles have a limited space that can be occupied by pickers. Cross aisle blocking occurs when a certain part of a cross aisle is occupied by another picker because of heavy traffic or an in-the-aisle blocking situation that diffuses into a cross aisle.
6. *Depot Blocking* – when two or more pickers want to use the depot at the same time. Usually, the depot is referred to as a location in the warehouse at which orders are consolidated. In the case of RFH, the buffer zone at which roll containers are stored after they leave the auction process can be considered as the depot.

At RFH, pick face blocking is the most prominent type of congestion. Some in-the-aisle interferences are observed, however this thesis focuses on pick face blocking as retrieving this from the data can be done relatively accurately.

The paper from Parikh (2006) is one of the first academic articles that focuses on blocking situations in wide-aisle order picking systems. This article identifies the effect of different picking strategies on congestion and concludes that a zone picking strategy results in the lowest amount of congestion for high levels of pick density, as is the case at RFH. Parikh & Meller (2009) concluded that when workers pick multiple products at a pick face, blocking increases monotonically with an increase in the pick-density (the probability of stopping and picking at a pick face). This shows that congestion increases when workers spend more time on picking than on travelling. Also, pick face blocking increases along with the number of workers in the system. Not surprisingly, blocking decreases as the size of the warehouse increases, due to the reduced pick-density as pickers spend relatively more time on traveling.

Next to the blocking situations defined in existing literature, interviews with managers at RFH and observations of the process have indicated that a pick face may be blocked by a queueing line. This indirect form of pick face blocking occurs when a worker that experiences pick face blocking, blocks another pick face in the aisle for one or multiple workers behind him. A queueing line may consist of up to six workers blocking multiple pick faces at the same time, which is a sincere problem for RFH.

3.3 Congestion Evaluation

Next to the type of blocking situations in order picking systems, Huber (2014) addressed the different methods that are most commonly used to analyze blocking situations. The analyzation of different type of blocking situations in academic research is mostly based on four methods:

1. *Simulation* – the imitation of processes that is used to compare order picking systems with different storage strategies. A variety of models has been developed by researchers to evaluate the effects of different order picking and storage strategies on congestion.
2. *Analytical methods* – probability models have been developed to calculate the decrease in throughput (order lines per hour) as a function of the number of possible blocking combinations and number of workers.
3. *Markov chains* – stochastic models that aim to determine the fraction of total time a picker is blocked.
4. *Queueing networks* – the study of waiting times that occur when a particular storage space is occupied by another order picker. Aimed to predict queue lengths and waiting times of systems with multiple interacting elements.

The abovementioned methods to analyze congestion in order picking systems are of a highly theoretical nature. Sandbrink's (2016) research on congestion in a real life warehouse provides a perfect example for this thesis research. Sandbrink (2016) also uses a data-driven approach in order to quantify congestion. Although he uses a different type of data (spatial-temporal) than what is available for this research, his research provides a perfect example for this thesis. Sandbrink (2016) concluded that congestion is nothing more than worker accumulation that leads to interactions and blocking situations. However, he focuses on solely visualizing and reducing congestion, instead of quantifying the amount of time lost due to congestion.

Beamon (1999) created a method to measure congestion in material handling systems through an index, based on the following assumptions:

- All material handling vehicles travel at the same speed
- Loaded vehicle speed equals empty vehicle speed
- Data can be collected at the beginning and end of each guide path link

The final assumption states that a computer can communicate with each vehicle at all control points in the system (being the start and end of a travel link). In order to compute each link's completion time, these two datapoints have to be available. Theoretical travel times that represent a congestion-free system are introduced to calculate the congestion index. Subsequently, the congestion index measure for each link *i* is computed through the following formula:

$$Congestion\ index_{\mathrm{i}} = 1 - \frac{theoretical\ travel\ time_{i}}{actual\ travel\ time_{i}}$$

Note that the congestion index is equal to zero when the actual travel time matches the theoretical travel time and that the congestion index approaches 1 for highly congested links.

3.4 Conclusion

Order picking systems have been widely discussed in existing literature, providing a strong theoretical establishment for the RFH case study. Also, order picking congestion has had a lot of attention from academic researchers. However, not a single paper on congestion deals with a reversed order picking system comparable to that of RFH. The same holds for the congestion evaluation methods, even though they are quite extensive, not a single example is directly applicable to the RFH case study. Therefore, the methods that are used for this thesis should not be solely focused on reproducing existing evaluation methods. They have to be able to take the differences between existing literature studies and the RFH case study into account.

Lastly, the overwhelming majority of literature on congestion is based on theoretical models, whereas this thesis aims at quantitative analysis of congestion. The two papers that used a quantitative method did not result in a quantification of congestion, they merely computed derived index numbers. This implies that new methods will have to be developed in order to quantify and explain congestion from real-life datasets.

CHAPTER 4 METHODOLOGY

This chapter describes the methods that were used to measure and explain the congestion in RFH's reversed order picking system. Section 4.1 explains the type of research that was used. Section 4.2 provides a thorough understanding of data mining and its application within this research. The procedure that was used to retrieve valuable information for congestion quantification is outlined in Section 4.3. Accordingly, the adjustments and expansion of the existing datasets are stated in Section 4.4. Based on the adjustments that were made, Section 4.5 describes the method that was used to compute actual travel times in the warehouse for each observation. Subsequently, with the new available data through actual travel times, some outliers were removed from the data as discussed in the sixth section of this chapter. Section 4.7 introduces the methods that were used to explain congestion. Finally, the last section lists the main conclusions of this chapter, pointing towards this thesis' added value to existing literature.

4.1 Methodological Approach

To answer the main research question of this thesis, a number of quantitative approaches were used. In Chapter 2 it was already mentioned that a lot of data is available from RFH's reversed order picking process. Since the aim of this thesis is to identify and quantify congestion in the system, a quantitative study was the most obvious manner to approach the problem. Through quantitative research, this thesis aims at creating contextual real-world insights in reversed order picking systems. An approach similar to that of Beamon (1999) was used to quantify congestion in RFH's distribution process. The most important input for this approach is the actual travel time between two subsequent locations. This information was gathered by combining the trace distributors-file and the voice logfiles.

All of the datasets regarding congestion quantification concerned Monday the 26th of August in 2019. Since Monday is the busiest weekday for RFH in terms of worker activity and the number of roll containers that go through the process, this day is known for a significant amount of congestion.

4.2 Data Mining

At the cornerstone of this thesis research lies the practice of data mining, "*the process of secondary analysis of large databases aimed at finding unsuspected relationships which are of interest or value to the database owners*" (Hand, 1998). Data mining is defined as a form of secondary analysis since it is very much an inductive way of data analysis and aims at retrieving relationships which are not apparent at first sight. Identifying these relationships, may result in two type of structures: models and patterns. Models are characterized as overall summaries of datasets. A pattern is "a local structure, possibly referring to only a relatively small number of objects" (Hand, Blunt, Kelly, & Adams, 2000). Regarding the RFH dataset with logged activity, one could consider the temporary presence of congestion at a location in the warehouse as a pattern. With the emphasis of data mining on purely locating these patterns, it served as the foundation for this research.

4.3 Data Manipulation

As part of the data mining process, the existing datasets discussed in Chapter 2 were manipulated using Rstudio software. Since the trace distributors-file contains all logistical transactions and related departure timestamps, it was used as starting point for data mining.

In order to reflect the actual route a worker has covered in the warehouse, the trace distributors-file had to be manipulated through a number of steps. To start with, all EENKOOP- and RESTKOOP-transactions were removed because workers did not have to transfer products at a location for these stops, making these observations irrelevant for congestion analysis. Secondly, most observations with an additional voice notification in the TaskStatus variable did not reflect a point in time at which a worker received a task to drive to a new location. Hence, except for the "Kar vol" notification (when a workers tells the system he has to replace a roll container at a location himself), all of these observations were removed. Lastly, the trace distributors-file contained multiple observations for a single location if a worker had to transfer flowers from different product classifications. In order to reflect reality, these observations were consolidated into one stop by accumulating the number of trays and saving the first timestamp.

To determine actual travel time between locations, the arrival timestamp had to be retrieved from the voice logfiles. Since these logfiles were stored in large and complicated text files, extensive data mining resulted in matching the right arrival timestamp with each observation from the trace distributors-file. Accordingly, the data was extended with a new column to specify the time between the departure and arrival timestamp, from now on referred to as travel time. Also, each observation was equipped with a variable that indicates the location of departure. This variable was of key importance since it was used to accumulate total driving distance between the two locations.

4.4 Data Extension

Unfortunately, none of the datasets contains information on travel distance. For this reason, all possible two-way combinations of warehouse locations were computed by accumulating cross aisle distance and within-the-aisle distances. These combinations were stored in a separate Excel file and linked to the trace distributors-file to compute total travel distance for each observation. Measurements were conducted to get standard dimensions within the aisles and on the cross aisle. Eventual cross aisle and within-the-aisle distances were based on these standard dimensions. A more detailed description of the determination of the travel distance can be found in Chapter 6.

4.5 Determination of Travel Times

Travel time is the difference between the departure and arrival timestamp. The departure timestamp is created when a worker receives the task to drive to a specific location in the warehouse. The arrival timestamp is created when a worker verifies he is at the right location by mentioning the check number, which is a random number placed on the ground at each storage location. This check number can only be mentioned when a worker physically arrives at a location, leaving no room for measurement errors. However, with respect to the departure timestamp the following assumption had to be made:

A1: *A worker finished transferring flowers and is only able to start his journey when he receives a new task.*

This assumption implies that the departure timestamp is the exact moment of departure. In other words, a worker would get on his electric truck when he gets his new assignment without any delays. To justify this assumption, RFH's voice-demonstration box was used. Using this box allowed for listening in into a worker's live voice communication. This was done for the distribution process of September 17 and showed that a worker is most likely to immediately step onto his truck when he receives a new task. However, for some cases it turned out that the worker got a new task and spent time on other activities before actually departing from a location. This includes work-related activities such as doing additional checks on quantities but also chatting with colleagues. Ultimately, this led to an upward bias in travel times since not all seconds of this derived variable were actually spent on driving. To be able to get valuable insights in congestion, actual travel times had to be compared with theoretical travel times. These theoretical travel times required trustworthiness in order to make a fair comparison between theoretical travel times and congested travel times. Hence, these standard travel times have been developed through a simple regression model, as will be described in Chapter 6.

4.6 Data Cleaning

After each observation was extended with travel time and distance, some remarkable outliers could be removed. To start with, 0.01% of the observations showed a negative travel speed as a result of an arrival timestamp that was created earlier than the relating departure timestamp. This impossible travel speed stemmed from a mistake in the way some voice logfiles were stored and could not be fixed by the R-script that was used. Since only a small fraction of data concerned this mistake, deleting them did not generate considerable problems for further analysis.

Secondly, 4.1% of the remaining observations that showed a travel speed higher than the maximum speed of 12 kilometers per hour (mechanically limited top speed of electric trucks) were removed. This speed could have been caused by the same mistake in the voice logfiles as was the case with negative travel speed. However, the departure and arrival timestamp may also be points in time that lay close to each other because workers were not using the voice picking system properly, as discussed in the previous section.

4.7 Statistical Analysis

The congestion patterns unveiled by the process of data mining gave valuable insights into RFH’s reversed order picking system. However, locating these patterns was not the single objective of this research. Statistical analysis in the form of regression modelling was conducted in order to explain why specific congestion patterns occurred in the data. This statistical analysis was performed using the latest Stata package, STATA 15 MP. Moreover, congestion was analysed aggregated for the whole warehouse per five-minute interval. Since the data from August 26 did not provide enough observations for decent analysis, all weekdays’ datasets from Week 35 were used. Various multiple regression models were considered to test whether the identified variables had explanatory power on the congestion patterns.

4.8 Conclusion

A lot of work was conducted before this thesis' actual research into congestion could start. The methods described in this chapter were developed specifically for the RFH case study and provided new insights to existing literature. This chapter described how the desired information on congestion was unravelled from the data. Retrieving the data required sophisticated R-scripts that none of the managers at RFH had been able to produce before the start of this thesis' research. The new information provided by the methods that were used, served as crucial input for this thesis.

CHAPTER 5 VISUALIZATION

This chapter is a first step in the analysis of RFH's reversed order picking system. A first glance at the data provides input for further analysis and is a crucial step to start the research. As mentioned in the previous chapter, all datasets provided by RFH for this research consider the 26th of August in 2019. Section 5.1 summarizes the key statistics concerning the trips on that day. Section 5.2 includes heatmaps of the number of transactions that are completed within the warehouse aisles. A new method to identify *pick face blockings* in the warehouse is outlined in Section 5.3. Subsequently, Section 5.4 introduces the type of voice notifications that relate to this thesis' research into congestion. Lastly, the number of active workers in the order picking system are reviewed in Section 5.4. Section 5.5 sums up the most important findings of this chapter.

5.1 Key Statistics

A total of 4.530 roll containers with at least one MEERKOOP-transaction went through the distribution process on the 26th of August. On average, it took the workers eight and a half minutes to distribute all the flowers from a single roll container. This is the total time between the moment at which a worker receives his first task for a specific roll container and when all trays are transferred to their designated storage locations in the warehouse. Note that this excludes the time spent on checking products on a roll container after it is picked up from the buffer zone and driving back to the buffer zone after all transactions on a roll container are distributed.

The workers had to transfer 19 trays with flowers, on average, per roll container. Some outliers of over 100 trays per roll container are observed. These are examples of very small boxes containing mostly Gerbera's, which are small and light flowers. The products were distributed to an average of 7 locations per roll container. In total 29.320 stops were made to transfer flowers.

5.2 Warehouse Heatmap

The flower auction in Naaldwijk is equipped with 8 clocks in total, each with a different product segment. All clocks start at six o'clock in the morning and have a specified order in which the different product groups are auctioned. As the buyers that are spread over the warehouse may have different buying patterns, the number of transactions per client location and aisle can be very volatile over a day. RFH follows the first-in-first-out principle concerning the buffer zone of flowers that have been sold at the auction clocks. This means that flowers that are auctioned first will be the first ones to enter the distribution process. Therefore, buyers that are active in the same product segment and have a storage location within the same aisle or even next to each other may cause substantially congested aisles.

Because the distribution process of RFH is such a dynamic process and congestion may appear only temporarily, the data is visualized on a five-minute interval. Each order line from the trace distributors-file contains a departure timestamp that is created when a worker completes a transaction at a storage location. These timestamps are rounded up to create aggregated data for the visualization of the process.

Figure 5.1 shows a heatmap of the number of transactions per aisle in the South section for the first two hours of the distribution process. From the conditional formatting that was used in the Microsoft Power BI tool it can be deducted that aisle 5 was the most crowded. Also, the number of transactions seems somewhat clustered. For instance, aisle 2 had more than 20 transactions per five minutes from 6:50 until 6:55 AM but significantly lower number of transactions before and after this timespan. This might be triggered by a specific client’s buying behavior.

Figure 5.2 shows that location 511 caused the most transactions in aisle 5. Moreover, the clustering of transactions holds for single locations as well. For instance, location 507 shows crowded timespans and periods of time without any transaction.

Number of stops per aisle

Timeframe	1	2	3	4	5	6	7	8	9	10	11	12
06:05:00	6	15	6	5	2	6	5	3			3	3
06:10:00	20	17	29	22	36	19	14	11	12	16	16	8
06:15:00	24	25	27	13	30	18	17	23	12	15	10	15
06:20:00	13	30	19	21	42	19	22	23	29	22	21	21
06:25:00	16	25	16	15	34	19	19	12	21	18	18	24
06:30:00	8	19	24	28	35	22	20	22	21	20	14	17
06:35:00	16	13	17	13	30	19	12	16	15	18	11	17
06:40:00	17	16	12	12	29	9	13	16	16	14	8	21
06:45:00	12	24	9	22	19	21	19	19	24	23	27	23
06:50:00	12	20	13	12	23	12	12	16	18	20	14	25
06:55:00	15	23	27	15	38	29	11	21	22	22	11	16
07:00:00	18	15	11	14	21	17	18	25	23	10	16	21
07:05:00	17	15	14	17	36	16	24	19	14	6	9	19
07:10:00	12	8	10	19	30	18	23	23	14	13	18	12
07:15:00	20	20	13	13	26	12	19	18	16	12	10	17
07:20:00	14	19	16	16	26	8	20	23	15	23	16	16
07:25:00	11	18	13	9	38	13	16	19	14	13	22	18
07:30:00	23	23	18	15	25	14	17	15	14	22	16	22
07:35:00	16	17	16	16	24	15	16	15	15	18	13	18
07:40:00	17	14	15	13	29	16	27	22	15	17	10	16
07:45:00	15	7	6	11	28	18	21	25	17	19	11	14
07:50:00	13	15	7	9	16	15	14	14	19	21	20	23
07:55:00	15	10	12	10	24	13	6	21	13	14	14	22
08:00:00	10	10	4	8	26	7	11	13	6	16	15	10

Figure 5.1 – Transactions per aisle

Number of stops per location

Timeframe	501	502	503	504	505	506	507	509	510	511	514	515	516	517	518	519	520	521	522	523
06:05:00			1								1									
06:10:00			3		1	2	4	4		6	6		3		2		4		1	
06:15:00		1	1	1	3	1	3	3		4	2		1		6	1		1	1	1
06:20:00		6	1	1	4	1	3	3	3	5	4		4		1	1	4			1
06:25:00		1			2	1	3			7	4		6		2	1	2		4	1
06:30:00		1	1		2			4	3	6	4		3		5	1	2	1	2	
06:35:00			1					2		9	7		3		2		4	1		1
06:40:00		1	2		4		3	1		5	2		7			1	2			1
06:45:00		1			1		3	1		3			1		4		1	1	2	1
06:50:00			1		2		3	2		3	2	1			2	1	4		2	
06:55:00					9	1	1	3		6	3		5		4		3	1	2	
07:00:00							4	4		4	2		1		2	1	1	1		1
07:05:00		3	1		4	2	1	3	1	6	1		4		3	1	4		2	
07:10:00		1			2	2	5	2	2	4	4		1		2	1	2		2	
07:15:00		2			5		1	1		2	2			1	4		6		2	
07:20:00			2		2	2	3	2		6	3		1		1		4			
07:25:00			2		4	3	2	3	1	3	3		4		6		3		3	1
07:30:00					3	1	4	1		2	4		1		1		6		2	
07:35:00			2			1	3	2	1	4	2		1		3		3		2	
07:40:00		1	1	1	4			1		6	1		1	1	3	2	4		3	
07:45:00					2	1	1	2	1	6	1	1	4	2	1	1	4	1		
07:50:00					1	1	3	2		2	1		1			1	3		1	
07:55:00				1	2	1	5	1		2	4		2				3		2	1
08:00:00					2	1	6		2	6	2		2			1	3		1	

Figure 5.2 – Transactions per location in aisle 5

As a result of this temporary increase in transactions within a specific aisle, congestion is likely to occur. However, there is no hard evidence on congestion from the number of transactions on itself as these might be well spread out over time.

5.3 Worker Interaction

According to managers at RFH, pick face blocking occurs quite regularly on Mondays. To test whether this is true and visualize congestion patterns in the process, pick face blocking situations were unravelled from the data. Pick face blocking implies that two workers want to access the same location at the same time, indicating that one worker is preventing another worker from doing his operations (Huber, 2014).

Checking whether a worker has experienced pick face blocking, can be done by specifying the time between the first worker's departure time and the second worker's arrival time at a specific location. With respect to pick face blocking, the following assumption was made:

A2: *A worker has potentially experienced pick face blocking if his arrival timestamp at location x is at most fifteen seconds later than the departure timestamp of the preceding worker that has visited location x.*

Within this assumption, a timespan of fifteen seconds is assigned for moving from the queueing line to the specific location[1]. Some observations showed that a worker's arrival timestamp at a location was created earlier than the preceding worker's departure timestamp. In practice, this turns out to be possible if a worker is able to mention the location's check number whilst queueing. Note that a worker could also just arrive at a location without any interaction, indicating that probably not all observations that arrive within the timespan of fifteen seconds have experienced pick face blocking. Also, other types of blocking, as for the next location in the aisle are not incorporated in this definition.

A significant amount of 5.395 observations experienced potential pick face blocking, which is more than 18% percent of all observations. When comparing the number of potential pick face blockings per five minutes with the total number of transactions through a heatmap, the patterns roughly align (Figure 5.3). Also, the pick face blockings occur mostly at the locations with the highest number of transactions in aisle 5 (Figure 5.4). This is illustrated by location 511, which had 297 stops and 62 potential pick face blockings, whilst location 502 had 48 stops and only 2 potential pick face blockings.

[1] Extensive research substantiating this parameter can be found in Appendix I.

Pick face blockings per aisle

Timeframe	1	2	3	4	5	6	7	8	9	10	11	12
06:05:00	3	5	2	0	1	0	2	0		0	1	0
06:10:00	4	1	4	1	13	4	3	3	3	1	2	0
06:15:00	3	5	6	2	7	2	4	3	2	1	3	1
06:20:00	3	6	4	7	10	3	3	2	8	7	5	2
06:25:00	3	4	6	7	9	3	3	3	5	2	2	7
06:30:00	1	3	3	5	10	5	2	2	6	7	2	1
06:35:00	2	0	4	4	7	6	4	2	4	8	2	2
06:40:00	2	0	2	5	11	1	1	4	6	3	2	2
06:45:00	0	2	0	4	2	5	4	2	2	6	7	4
06:50:00	4	8	1	0	3	3	1	6	5	5	3	7
06:55:00	4	0	6	1	16	6	1	3	6	4	0	0
07:00:00	4	2	4	0	5	1	6	5	3	1	1	3
07:05:00	2	0	2	1	9	8	2	5	3	3	0	1
07:10:00	3	0	0	3	5	4	4	1	3	1	1	2
07:15:00	6	1	2	2	7	2	4	5	4	3	0	1
07:20:00	1	1	5	8	8	0	1	4	1	5	6	3
07:25:00	6	1	5	1	10	2	3	3	3	3	9	2
07:30:00	3	2	5	4	6	2	5	1	3	5	2	1
07:35:00	7	3	2	4	6	0	9	3	1	2	3	3
07:40:00	0	0	7	0	11	1	7	6	4	6	2	2
07:45:00	1	0	0	2	5	3	3	4	3	6	2	0
07:50:00	3	2	0	2	4	3	1	4	3	4	7	5
07:55:00	1	0	3	3	4	1	1	4	2	2	5	2
08:00:00	1	0	0	1	5	0	2	0	1	5	2	0

Figure 5.3 – Pick face blockings per aisle

Pick face blockings per location

Timeframe	501	502	503	504	505	506	507	509	510	511	514	515	516	517	518	519	520	521	522	523
06:05:00			1				0			0	0									
06:10:00			0		0	1	0	2		1	2		1		3		3		0	
06:15:00		1	0	0	2	0	1	0		1	0		0		2	0		0	0	0
06:20:00		0	0		1	0	0	0	0	3	3		1		0	0	1		1	0
06:25:00		0			0	0	0	0		4	0		4		0	0	1		0	0
06:30:00		0	0					0	1	2	4		1		2	0	0	0	0	
06:35:00			0					0		4	1		0		0		2			0
06:40:00		0	1		2		2	0		3	0		3		0	0	0	0		0
06:45:00		0			0		1	0		0					1		0		0	0
06:50:00			0		0		0	0		0	0	0			0	0	2		1	
06:55:00					4	0	1	2		3	1		1		2	0	1	0	1	
07:00:00		0			0		1	1		2	1		0		0		0	0		0
07:05:00		0	0		0	0	0	0	0	5	0		1		1	0	2		0	
07:10:00		0			0	0	3	0	0	0	0		0	0	1		1		0	
07:15:00		0	0		1	0	0	0		0	0		0		1		5		0	
07:20:00			0		0	0	1	0	0	2	4		0		1		0		0	
07:25:00			0		1	1	1	1		0	0		2		1		2		1	0
07:30:00			0		1	0	1	0	0	0	2		0		0		1		1	
07:35:00					0	0	0	0		1	1				2		2		0	
07:40:00		0	0	0	2		0	0		4	0		2	0	1	0	1		1	
07:45:00					0	0	0	1	0	2	0	0	0	1	0	0	1	0		
07:50:00				0	0		3	0		0	0		0			0	1		0	
07:55:00					0	0	2	0		1	0		0				1		0	0
08:00:00			0		0		2		1	2	0					0	0		0	

Figure 5.4 – Pick face blockings per location in aisle 5

5.4 Voice Notifications

As mentioned in Chapter 2, the TaskStatus variable contains information with the type of notification that a worker has sent to the voice picking system. The following two type of notifications are relevant for congestion analysis:

- *Pad overgeslagen*: when a worker decides to skip a whole aisle;
- *Locatie overgeslagen*: when a worker decides to skip a specific location.

The *locatie overgeslagen* notification is a possible data polluter as a single worker may use it more than once for the same aisle. For example, when a worker thinks an aisle is too congested to enter, he might send this notification to the voice system. However, if he has to visit more than one storage location in the same aisle, the system will re-direct the worker into the aisle. Consequently, the worker will send a second notification to the system since he still thinks the aisle is too crowded, creating an upward bias in the number of notifications for a specific aisle.

To overcome this upward bias, the aggregated number of the two notifications per five minutes has to consist of unique roll container numbers. By means of this, the aggregated dataset at a five-minute interval will count multiple *locatie overgeslagen* notifications by the same worker within an aisle as one instead of summing them up.

As might be expected from the number of transactions and potential pick face blockings per aisle, Figure 5.5 shows that aisle 5 has had the most notifications on both aisle and location skipping. Although overall patterns of voice notifications and the number of transactions per aisle show similarities, they do not perfectly align. The extreme number of six notifications in aisle 5 at 7:15 AM does not correspond with the highest number of transactions. Note that this may indicate that not all workers actually use the voice notifications to skip aisles or locations. They are likely to get in line and wait until a location is free for utilization.

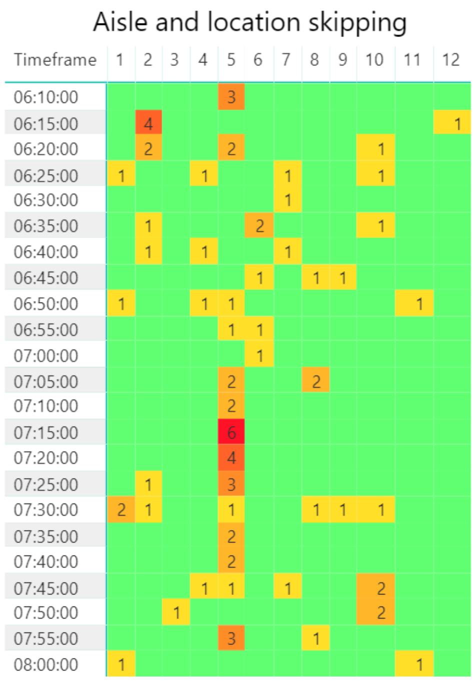

Figure 5.5 – Voice notifications per aisle

5.5 Active Workers

The number of workers that are active in the distribution process is logged by RFH through a file that counts the unique employee numbers that complete a transaction per fifteen minutes. This measurement is based on the assumption that a worker will never need more than fifteen minutes to drive to an aisle and transfer the required number of trays. Figure 5.6 depicts the number of active workers in the system and clearly shows that the workers take breaks in shifts from 8:00 until 9:30 AM. The drop in the number of potential pick face blockings seems to align with this period of time (Figure 5.7).

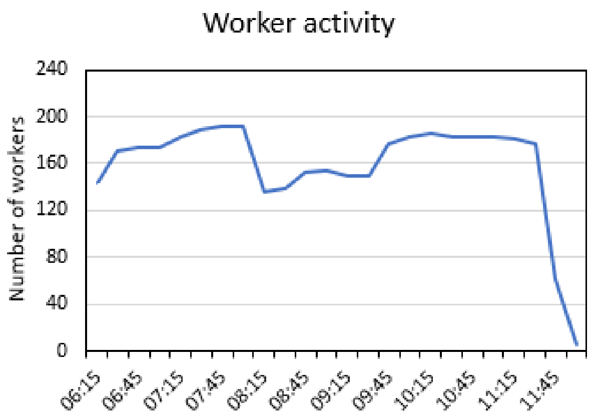

Figure 5.6 – Number of active workers per fifteen minutes

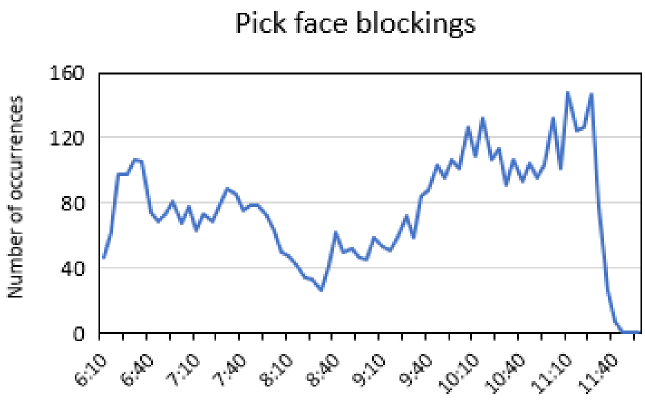

Figure 5.7 – Number of pick face blockings per fifteen minutes

5.6 Conclusion

The visualization methods that have been used in this chapter show that RFH's reversed order picking system is highly volatile in terms of the number of transactions per aisle and location within an aisle. Correspondingly, the number of potential pick face blockings fluctuates over time. With the identification of these potential pick face blockings, this chapter has established a new tool for RFH to measure and visualize congestion. Moreover, this method of congestion visualization adds new insights to the existing literature on congestion in order picking systems.

CHAPTER 6 CONGESTION QUANTIFICATION

As there is no concrete measurement of the time lost due to congestion, it has to be derived from the data by specifying theoretical travel times and subtracting these from actual travel times. This chapter develops a method that enables congestion quantification in RFH's system. An approach similar to that of Beamon (1999) is used, based on the idea that congestion delays the moment of arrival at a location, thus extends actual travel time. This approach uses index numbers to indicate whether travel links are congested or not. Beamon's (1999) approach is extended to allow for quantification of congestion instead of merely creating index numbers.

Section 6.1 starts with an introduction of the assumptions and variables that are used to set up theoretical travel times. Based on the findings in this section, Section 6.2 introduces a method that calculates theoretical travel time for each observation. Through comparison of theoretical and actual travel times and controlling for observations with potential pick face blocking, Section 6.3 quantifies congestion in the system. Finally, Section 6.4 concludes the most important findings of this chapter.

6.1 Theoretical Travel Time and Pick Face Blockings

The travel distances covered by workers and related travel times are based on the following assumptions:

A3: *All workers follow the compulsory route without taking illegal shortcuts.*

A4: *All vehicles travel at the same, constant speed.*

A5: *Within-the-aisle travel speed is lower than cross aisle travel speed.*

Beamon (1999) determined theoretical congestion-free travel times based on theoretical travel speed. However, theoretical travel speed may not reflect reality as it does not take the properties of a warehouse into account. To approach congestion-free travel times, observations from the first ten minutes of the distribution process of Thursday the 29th of August are used. Since Thursday has the lowest worker activity, total number of transactions and potential pick face blockings in the first ten minutes of the process, using these observations as input for theoretical travel speed is the most preferred option. More importantly, the fraction of potential pick face blockings was only 10.1% on Thursday compared to 14.8% of all transactions on Monday (Table 6.1).

Table 6.1 – Key statistics from first ten minutes of the distribution process in Week 35

	Monday	Tuesday	Wednesday	Thursday	Friday
Active workers	95	93	89	73	84
No. of transactions	372	335	357	308	363
No. of pick face blockings	55	42	52	31	48
Percentage of pick face blockings	14.8%	12.5%	14.6%	10.1%	13.2%

Only a small fraction of this subset of data (but still 10%) contains congestion because not all workers start at the same time and the chance of pick-face blocking is significantly smaller than at a later stage of the process. At RFH, there is no occurrence of a "moving cluster" as a result of temporal clustering of pick tour starts (Sandbrink, 2016). Moreover, the number of delayed observations due to potential pick face blocking was significantly lower than during the rest of the process as can be seen in Figure 6.1. The second lowest value for potential pick face blockings per ten minutes was already 69, compared to 31 from the first ten minutes of the process.

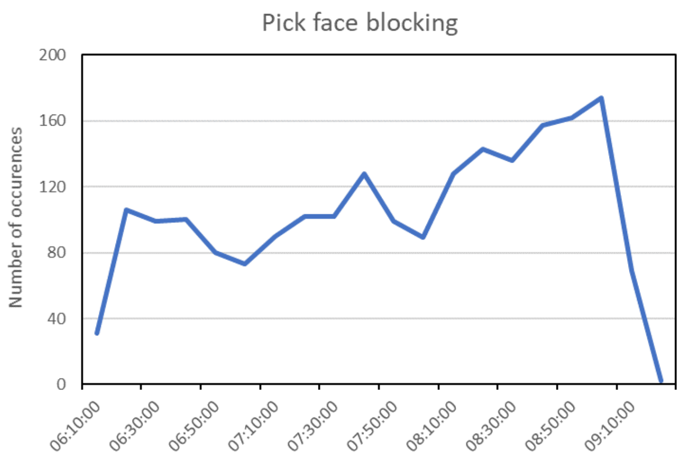

Figure 6.1 – Number of pick face blockings per ten minutes on August 29

6.2 Theoretical Travel Time Modelling

Supported by the findings from the previous section, a simple regression model is set up in order to compute real-life theoretical travel times. The data that is used for this model contains 308 observations from the first ten minutes of the distribution process on August 29th , as described in the previous section and the following simple regression model is used to predict congestion-free travel times in seconds:

$$\texttt{Travel\_time} = \beta_0 + \beta_1 \texttt{Travel\_distance} \tag{6.1}$$

After deleting the 31 observations that potentially experienced pick face blocking, which should not be included in theoretical travel times as they prolong travel time due to interaction, a scatter plot is made to test whether the data was suitable for regression (Figure 6.2).

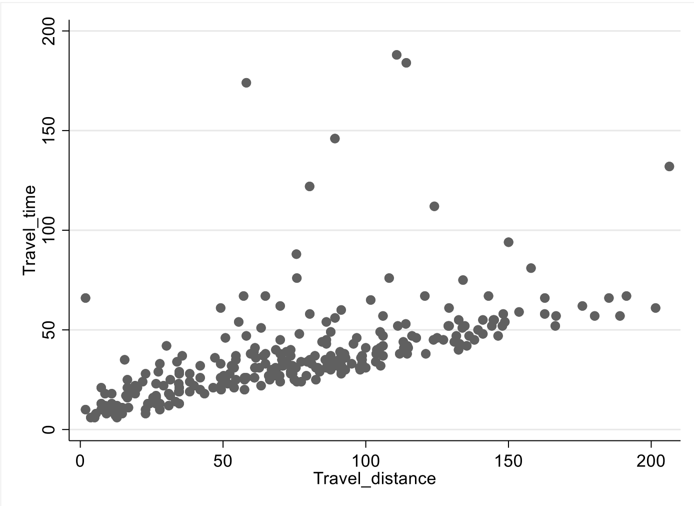

Figure 6.2 – Scatterplot of travel time with outliers (subset August 29th)

The plot shows a clear band of data with a width of approximately 20 seconds for all values of `Travel_distance`. The lower bound resembles observations with the highest travel speed of 12 kilometers per hour (established after outlier removal as described in this thesis' Methodology). The upper bound lies 20 seconds above this lower bound. However, some outliers with higher values for `Travel_time` are observed outside this thick band of observations. After removing observations outside the band of data from the sample (18% of the observations in this sample), 228 congestion-free observations remain as input for the regression model to compute theoretical travel times. As can be seen in Figure 6.3, these observation follow a clear linear trend perfectly suitable for regression modelling through Ordinary Least Squares (OLS) estimation.

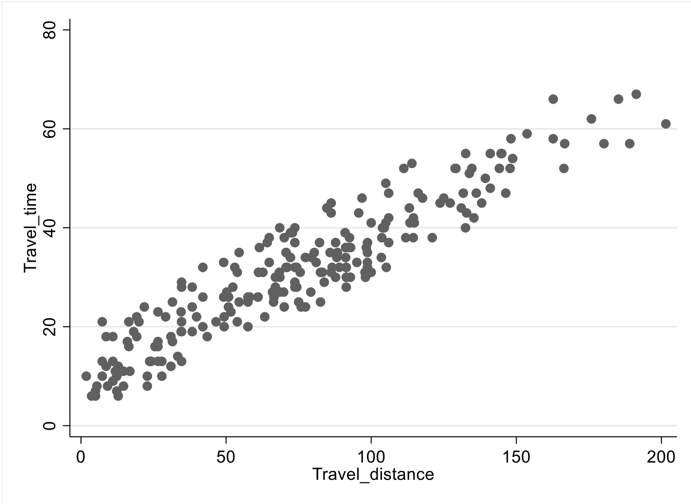

Figure 6.3 – Scatterplot without outliers (subset August 29th)

Through simple linear regression the following coefficients are found:

- β_0 = 9.878 seconds
- β_1 = 0.287 seconds per meter

The regression output indicates that travel distance has a significant effect at the 5% level on travel time; one meter of travel distance increases travel time by 0.287 seconds, on average. The model has a an R-squared of 0.873, a detailed discussion on the OLS assumptions of the regression model that was used to predict theoretical travel times can be found in Appendix II.

As discussed in Appendix II, the electric trucks that are used by the workers have different characteristics in terms of top speed and acceleration. This explains why the data in Figure 6.3 is not normally distributed around a mean but shows a rather scattered pattern. Therefore, to set up fair theoretical travel times and account for this difference in truck characteristics, the upper bound of the data in Figure 6.3 was used for identifying congestion by adding ten seconds to the constant such that β_0 = 19.878 seconds. This translates into the following formula that is used to predict theoretical travel times:

$$\texttt{Travel\_time} = 19.878 + 0.287 * \texttt{Travel\_distance} \qquad (6.2)$$

6.3 Congestion Quantification

The following variables are used to measure the congestion in RFH's reversed order picking system:

A_i = actual total travel time for observation i in $seconds$

T_i = theoretical travel time for observation i in $seconds$

D_i = distance covered for observation i in $meters$

P_i = dummy that takes the value 1 if observation i experienced pick face blocking

C_i = congestion for observation i in $seconds$

N = total number of observations

i = observation number

Once the theoretical travel time for an observation is determined, the amount of time wasted due to congestion is easily quantified:

$$C_i = A_i - T_i(D_i) \tag{6.3}$$

Note that the actual travel times are inferred from the two timestamps that are available in the manipulated dataset and the function $T_i(D_i)$ equals the regression output from the previous section (formula 6.2) with the adjusted β_0. Observations that did not experience potential pick face blocking but exceeded theoretical travel times may have had other issues that should not be recognized as congestion. For example, chatting with colleagues or asking a tutor for help prolongs travel time but should not be identified as congestion.

6.3.1 Results

The results shown below that are calculated through formula 6.3 are based on the data of August 26 and exclude observations with a travel speed higher than 12 kilometers per hour (deleted as outliers). Moreover, five observations with more than 10 minutes of congestion were deleted since they are not likely to stem from interferences such as pick face blocking. Table 6.2 shows the key statistics from the congestion quantification for observations with and without potential pick face blocking (i.e. P_i = 1 and P_i = 0).

Table 6.2 – Congestion quantification for August 26

	P_i = 1	P_i = 0
Number of observations	5.395	23.828
Total congestion in seconds	124.679	83.160
Average congestion per observation in seconds (exceeding theoretical travel time)	23,11	3,49
Total excessive manhours	34,63	23,10

As can be seen from Table 6.2, nearly 35 manhours were wasted due to congestion on the 26th of August. Additionally, more than 23 manhours were wasted by observations that did not experience pick face blocking. As mentioned earlier, these observations could have been delayed by various causes which are hard to subtract from the data.

There is a noticeable difference between the average time lost for the two values of P_i. On average, observations that experienced pick face blocking lost more than six times as much time as observations without potential pick face blocking. The difference between the average congestion per observation for the two values of P_i can be considered as the true waste caused by pick face blocking, which is 19,62 seconds. Multiplying this with the number of observations that experienced pick face blocking gives the true time lost due to congestion: 29 hours and 24 minutes.

Surprisingly, 1.789 observations that are flagged by the potential pick face blocking dummy did not exceed their theoretical travel time. This could be the result of the differing characteristics per electric truck. Also, the fact that the regression line for theoretical travel times was moved upwards by 10 seconds could have caused the lack of congestion for these observations. Lastly, it is not unlikely to assume that these observations did not interact at all with the preceding observation at the same location. If a worker arrived at a location fifteen seconds after his predecessor left, there does not necessarily have to be interaction.

Comparing the results with the *waiting-time paradox*, which states that "a passenger arriving at a bus stop will probably have to wait considerably longer than about half the interarrival time, X say, of two buses" (van Harn & Steutel, 1995), indicates that the same holds for a worker that has to wait for a colleague whom is blocking a location. With the average transfer time of 26 seconds (i.e. the time that a worker blocks a location for someone else), one would expect the average waiting time for workers that experience pick face blocking to be half of this, namely 13 seconds. However, observations with potential pick face blocking show an average waiting time of nearly 20 seconds. For exponential distributions, the waiting-time paradox states that the average waiting time would equal the average interarrival time. Even though this does not hold for the waiting time in RFH's system, it appears that the average waiting time is considerably longer than half the transfer time, as the paradox suggests. At the same time this calculation shows that the number obtained as average waiting time, is a realistic value.

6.4 Conclusion

This chapter has provided a mechanism that is able to quantify congestion in terms of manhours lost due to pick face blocking in RFH's system. The method of Beamon (1999) was extended to allow for this quantification. The results have shown that potential pick face blockings prolong travel times on average with almost 20 seconds, compared to situations without interactions. In other words, RFH lost 30 manhours of waiting time on August 26. Further investigation into the time that is wasted on all other days in the year should indicate how much money is drained due to pick face blockings.

CHAPTER 7 IDENTIFICATION

This chapter identifies possible explanatory variables of the number of potential pick face blockings in RFH's reversed order picking system. Section 7.1 elaborates on the type of model that is developed to analyse the amount of congestion in RFH's system. Thereafter, the possible causes of congestion are identified in Section 7.2. The last section lists the most important takeaways from this chapter.

7.1 Modelling

Since not all observations with pick face blocking lose time, as shown in the previous chapter, this thesis focuses on explaining congestion in the form of the number of potential pick face blockings instead. Moreover, the congestion that has been quantified may be a result of other issues than pick face blocking. Events such as chatting with colleagues and asking for assistance from an instructor are hard to explain with data and result in an upward bias in congestion which cannot be corrected for. Alternatively, the number of potential pick face blockings is a direct result of the different parameters of an order picking system, which is why it is better suitable as dependent variable in regression modelling.

As mentioned by Sandbrink (2016), congestion occurs at a certain place and a certain moment in time. Consequently, explaining the occurrence of individual congestion is very challenging and complicated. Therefore, this chapter aims at explaining congestion at an aggregate level, being RFH's complete warehouse. The approach is more precise compared to the approach from Sandbrink (2016), as congestion is explained at a five-minute interval. Since congestion is explained at an aggregate level, data from August 26 does not provide sufficient observations for decent analysis. To overcome this problem, all datasets from Week 35 are used (Monday until Friday) to explain the number of pick face blockings at a five-minute interval (`n_pfb`).

7.2 Pick Face Blocking Causes

Pick face blocking occurs when multiple workers want to transfer flowers at the same location at the same moment in time. The probability of multiple workers wanting to access the same location heavily depends on the amount of time that workers use to transfer products compared to travel times (Parikh & Meller, 2009). The more time workers spend on transferring products, the more likely it is that they meet each other at the same location in the warehouse. On its turn, the ratio of time spent on transferring products to time spent on traveling depends on various factors. Variables that influence this ratio aggregated for the complete warehouse at a five-minute interval are listed in this section.

7.2.1 Number of Transactions

The explanatory variable `n_stops` denotes the number of stops that are made within each five-minute interval. Logically, the chance of potential pick face blocking increases when more locations are being accessed for a given time interval. Figure 7.1 shows a scatterplot of values for the `n_pfb` and `n_stops` variables for all observations in Week 35. The obvious pattern shown by the plot is supported by a correlation coefficient of 0.908 between the two variables. The plot even seems to have a convex pattern, which is why this variable will also be tested with a polynomial. As can be seen from Figure 7.1, this variable's range runs from zero to roughly 550 stops made in a five-minute interval.

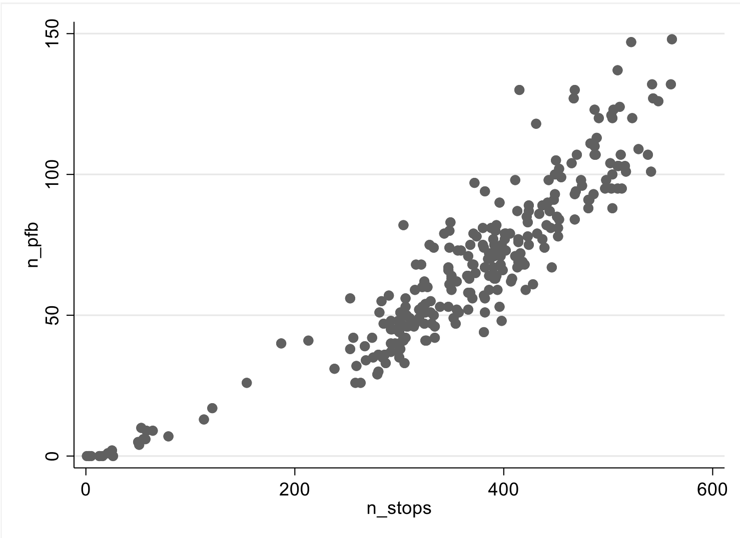

Figure 7.1 – Scatterplot of number of stops and pick face blockings per five minutes

7.2.2 Number of Workers

Next to the fact that the number of stops that are made in the warehouse affects potential pick face blockings, the number of workers that actually makes these stops influences the dependent variable. To give an example, if the same number of stops is made by more workers, there is a higher chance that two workers try to access the same location at the same time. Stated differently, the regression model should control for the number of workers in the system (variable `n_workers`). The number of workers in the system has proven to be one of the foremost influencers of congestion in order picking systems (Huber, 2014; Parikh & Meller, 2009; Sandbrink, 2016), which is why it cannot be left out in this thesis' regression model.

Figure 7.2 shows the number of unique workers that transferred products per five-minute interval. The end of each day is clearly visible as the number of workers drops to zero. The graph also shows a drop in worker activity as a result of the lunch breaks, except for Thursday. Since the process is relatively short on Thursdays, RFH often decides to work on without taking a break. As can be seen from Figure 7.2, the number of workers per five-minute interval ranges from zero to 175.

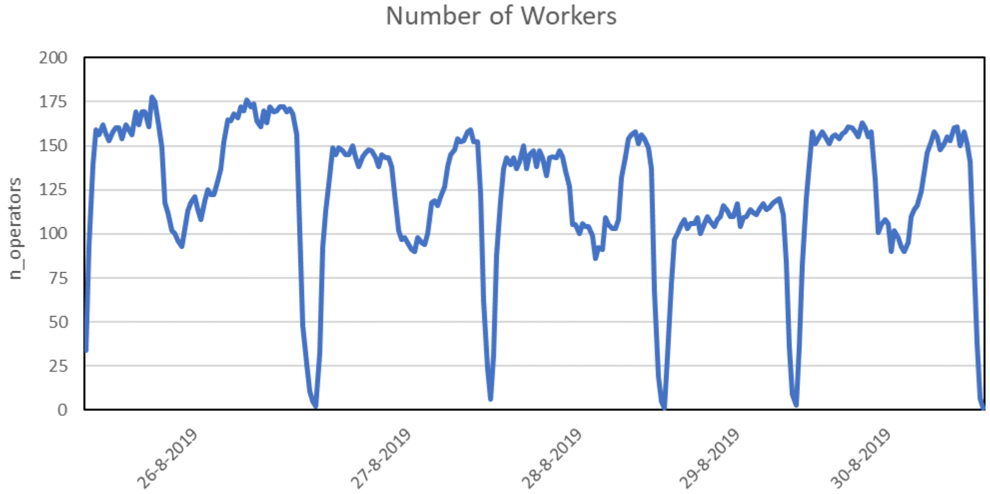

Figure 7.2 – Number of unique workers that transferred products per five-minute interval

7.2.3 Number of Locations

Another variable that has to be controlled for when explaining the number of potential pick face blockings is the number of locations that was accessed. To start with, on Monday all aisles and locations are utilized. On days with lower total throughput, some locations may not be used at all. Also, some aisles may be closed on these days. Besides the fact that some locations may be closed, the `n_locations` variable accounts for the number of locations that are actually accessed. For example, when only ten out of twenty locations in an aisle are visited, there is a bigger chance of congestion if the number of transactions is high compared to when all locations are visited. In a way, this variable indicates the size of the picking area per five-minute interval, which is a relevant factor for congestion in order picking systems (Parikh & Meller, 2009). This variable's range runs from 1 until 268 locations that were accessed within a five-minute interval.

7.2.4 Transfer Time

The three variables that have been discussed this far describe the process for a given five-minute interval. However, the amount of time spent on transferring products by workers might also explain a part of the variance in the number of potential pick face blockings. Specifically, the more time is spent on transferring products per five-minute interval, thus utilizing a location, the bigger the chance of pick face blocking within that specific time interval. Therefore, the model has to control for the time spent on blocking locations when explaining the number of potential pick face blockings through total number of transactions. The number of products that was transferred at a location has also been proven to be a significant influencer of congestion (Parikh & Meller, 2009). Nonetheless, the total amount of time spent on transferring will be even more powerful when explaining congestion since it accounts for the number of products and all other situations that induce a longer location utilization time. Examples of these situations are replacing full roll containers at a location and chatting with colleagues (as described in Chapter 2).

Total transfer time is also preferred over total number of products since it has a higher correlation coefficient with pick face blockings at the five-minute interval; respectively 0.845 and 0.813. Even though there is only a slight difference in correlation with the dependent variable, `transfer_time` per five-minute interval will be used as explanatory variable instead of the number of products.

7.2.5 Worker Experience

Through interviews with managers at RFH it has become apparent that employment turnover rates are high, resulting in a lot of new workers in the reversed order picking system. New workers require introduction and have to learn in order to become familiar with the process. "Worker learning in order picking is highly relevant in practice" and "most often observed in the time that is needed for searching for items" (Grosse & Glock, 2015). Therefore, new workers will require more time at a location to transfer products as they are not yet familiar with the operations. Thus, having a lot of new workers in the system might increase the possibility of congestion. To test whether the number of new workers in the process has influence on the amount of potential pick face blockings, the variable `beginners` is created. This variable accumulates the number of workers per five-minute interval with less than a total of 150 working manhours at RFH. This limit is set by RFH and is considered as the point at which a worker does not require active guidance anymore.

7.3 Conclusion

This chapter has identified five possible causes of congestion in RFH's reversed order picking system, both based on literature and quantitative analysis: the number of stops and a related squared term, the number of workers, total transfer time and the number of beginners in the system. Moreover, it started with an introduction into the model that will be used to explain congestion in the next chapter. The method that will be used is similar to that of Sandbrink (2016), however, by explaining congestion at a five-minute interval it uses a more precise approach.

CHAPTER 8 REGRESSION

This chapter tests whether the variables identified as possible causes of congestion in the previous chapter have explanatory power when including them in a multiple regression model. The dependent variable is `n_pfb`, the number of potential pick face blockings in the warehouse per five-minute interval. The regression model developed in this chapter serves as a tool that can be used to evaluate the distribution process at RFH. It enables managers to compare the actual amount of pick face blockings with a predicted number from the tool. Furthermore, they can relate this to possible drops in worker productivity, which is an actively monitored key performance indicator at RFH.

The first three sections of this chapter introduce the three different type of models that are tested in order to explain congestion. Section 8.4 gives an extensive review of the output from each model. Finally, Section 8.5 concludes on the best model that RFH can use to evaluate congestion.

8.1 Model I

Regression equation 8.1 describes model I, which tests for a linear trend between the number of potential pick face blockings and the total number of stops. It is assumed that the number of potential pick face blockings increases with β^1 for each additional stop that is made in the process.

$$\texttt{n\_pfb} = \beta^0 + \beta^1 * \texttt{n\_stops} \tag{8.1}$$

8.2 Model II

The second model that is used to explain the number of potential pick face blockings includes both `n_stops` and the squared value of the number of stops (`n_stops_sq`). This variable is added to test whether the slope of the regression line from model I increases along with the values of `n_stops`, resulting in equation 8.2.

$$\texttt{n\_pfb} = \beta^0 + \beta^1 * \texttt{n\_stops} + \beta^2 * \texttt{n\_stops}^2 \tag{8.2}$$

8.3 Model III

The third and last model is set up by using the forward stepwise command in Stata[2]. By using this command, Stata adds the most significant variable to the model until the next variable that is added, does not add any explanatory power to the model. The variables are selected based on their contribution to the R-squared of the model and individual t-statistics. The following variables are tested on explanatory power with respect to potential pick face blockings through the forward stepwise method:

- `n_stops` (number of stops made in the warehouse)
- `n_stops_sq`
- `n_workers`
- `n_locations` (number of locations over which the stops are distributed)
- `transfer_time` (in seconds)
- `beginners`

Model III tests for the best control variables that should be included when explaining the number of potential pick face blockings at the five-minute interval.

8.4 Regression Output Evaluation

All of the models that are discussed are fitted through Ordinary Least Squares (OLS) estimation. For each model, the underlying OLS assumptions are discussed in Appendix III. Only the motivation for model II, which is based on a violation of an important OLS assumption, is discussed in this section. The models are evaluated through their R-squared and F-statistic.

The coefficients that are found after creating the three models can be found in Table 8.1. This section discusses the parameters for each model. Note that this section does not interpret the constant of any of the models since this is not meaningful. Interpreting the constant would mean that no stops are made in the process, which is not observed and not interesting to look at either.

[2] The forward stepwise command includes variables with a p-value lower than 0.05 for the t-statistic.

Table 8.1 – Regression output.

VARIABLES	(1) Model I	(2) Model II	(3) Model III
n_stops	0.232***	0.0477**	0.434***
	(34.67)	(2.467)	(25.21)
transfer_time			0.00328***
			(7.32)
n_locations			-0.669***
			(-21.39)
n_stops_sq		0.000309***	
		(9.974)	
Constant	-19.57***	1.490	5.454***
	(-7.374)	(0.480)	(4.07)
Observations	274	274	274
Adjusted R-squared	0.815	0.864	0.938
F-statistic (p-value)		0.000	0.000

t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

8.4.1 Model I

The first model already shows an unquestionable relationship between the number of potential pick face blockings and the total number of stops that were made in the warehouse per five-minute interval. Total number of stops has a positive and significant effect at the 5% level on the number of potential pick face blockings; on average, one additional stop increases pick face blockings by 0.232. The model has a noticeably high R-squared of 0.815 indicating that 81.5% of the variability in total number of potential pick face blockings per five minutes can be explained by the changes in total number of stops that has been made in the warehouse.

8.4.2 Model II

Checking model I for violations of the OLS assumptions (as discussed in Appendix III) shows that the Zero Conditional Mean assumption is violated. Figure 8.1 shows that the error term does not show an expected mean of zero for all values of `n_stops`. Therefore, the second model was included with a polynomial of `n_stops`.

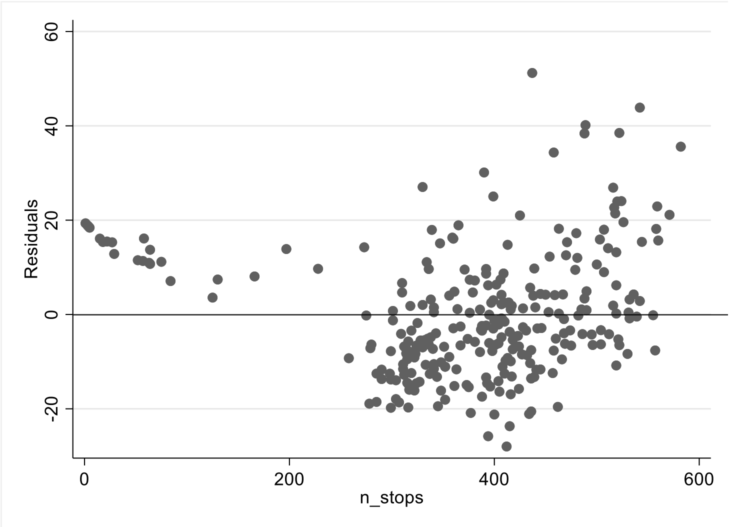

Figure 8.1 – Zero Conditional Mean violation model I

Table 8.1 shows that the F-statistic on joint significance of the explanatory variables in model II points out that including the quadratic term is statistically correct (p = 0.000). Also, the explanatory power of the model has increased to an adjusted R-squared of 0.864. So when accounting for the addition of the squared term, the models still explains more of the variability in potential pick face blockings.

The best way to interpret the regression coefficients of the model with polynomial is taking the derivative of the regression equation with respect to the explanatory variable `n_stops`. This derivative shows that the increase in number of potential pick face blockings depends on the initial value of the number of stops. Strictly speaking, the increase in potential pick face blockings is calculated as follows: 0.048 + 0.0006 * `n_stops`.

The difference in the slope of the regression curve can best be explained by giving examples for two meaningful values of `n_stops`; 300 and 500. For each of these values, an additional stop per five-minute interval increases the number of potential pick face blockings with respectively 0.228 and 0.348. Thus, a higher initial number of stops induces a larger increase in the number of potential pick face blockings. Which is higher than Model I, which predicts a constant increase of 0.232.

8.4.3 Model III

In the third model, all variables identified in the previous chapter are tested for their explanatory power on the number of potential pick face blockings in the process. As can be seen in Table 8.1, the forward stepwise method selects only three variables. The number of stops, locations and total transfer time form the model with the highest explanatory power.

The number of stops has a positive and significant effect at the 5% level on the number of potential pick face blockings; on average, one stop increases the number of potential pick face blockings with 0.434, keeping all other variables equal.

In addition, total transfer time also has a positive and significant effect (at 5%) on the number of potential pick face blockings. On average, each extra second spent on transferring products per five-minute interval increases the number of potential pick face blockings with 0.003, all other variables held constant. This variable seems to have a small impact, however, its range runs from roughly 2.000 to 15.000 seconds. So when comparing two five-minute intervals with a difference of 5.000 seconds in total transfer time it seems that, on average, fifteen more potential pick face blockings occur, ceteris paribus.

Lastly, the number of locations that are accessed has a negative and significant effect at the 5% level on the number of potential pick face blockings. Utilizing an additional location reduces the number of potential pick face blockings with 0.669 on average, ceteris paribus. This shows that the distribution of the number of stops over the warehouse matters. If the stops are made at fewer locations, more potential pick face blockings will occur. In a sense, this variable accounts for the spatial clustering of worker activity in the warehouse.

The third model has a joint significance on the number of potential pick face blockings ($p = 0.000$ for F-statistic) and an adjusted R-squared of 0.938. Thus, almost 94% of the variability in the number of potential pick face blockings can be explained by model III. Surprisingly, this model excludes the polynomial of `n_stops`, indicating that the violation of the Zero Conditional Mean assumption in model I is a result of Omitted Variable Bias instead of a misspecification of the model's functional form as the parameter for the number of stops changes compared to Model I.

It is also striking that the number of workers is not added to the model even though this is identified as the most important cause of congestion in existing literature. Apparently, the number of workers is too correlated with the other variables in the model to add any explanatory power. This is perfectly intuitive since workers act as a kind of mediator that enables a certain amount of stops to be made. Since its high correlation with the number of stops which has more explanatory power on potential pick face blockings, there is no need to include it in the model.

8.5 Conclusion

This chapter has shown that the number of potential pick face blockings in the process is perfectly suitable for regression modelling. A tool has been created that RFH can use to compare the actual pick face blockings per five minutes with. Moreover, it turned out that the number of stops, locations and total transfer time are the most important causes of congestion in the system. With these variables, almost 94% of the variability in the number of potential pick face blockings per five-minute interval can be explained.

The model developed in this chapter gives managers at RFH direction for further investigation into the reduction of congestion, requiring a good understanding of the explanatory variables. Figuring out how these variables can be influenced, gives RFH the power to control congestion in the reversed order picking system.

CHAPTER 9 CONCLUSION

This research examined congestion in the reversed order picking system of Royal FloraHolland. Through four sub-objectives the following research question was investigated: *Where does congestion occur in the reversed order picking system of Royal FloraHolland and what are its most important causes?* This chapter produces a proper answer to this question by summarizing the work that was done in order to achieve all sub-objectives.

9.1 Visualization

Design of a tool that is able to unveil pick face blocking from existing datasets.

Pick face blocking occurs when two or more workers want to access the same location in the warehouse at the same point in time, resulting in congestion through waiting times. Identification and visualization of these pick face blockings through heatmaps (page 28) showed that Royal FloraHolland's reversed order picking system is characterized by an uneven distribution of activity over time and throughout the warehouse. The occurrence of potential pick face blocking identified by the tool shows a pattern similar to the number of stops made within aisles and at locations. Moreover, the number of active workers and voice notifications on aisle skipping are related to potential pick face blockings.

9.2 Quantification

Extending an existing method that compares congested travel times with theoretical travel times.

The amount of time lost due to congestion in the system was quantified by extending Beamon's (1999) existing approach on congestion quantification (page 33). Theoretical travel times were set up by using data without congestion. After comparing these with actual travel times, it was demonstrated that potential pick face blocking prolongs travel time, on average, with almost 20 seconds. With respect to the distribution process of Monday August 26, this means that Royal FloraHolland spent 30 manhours on waiting time. The method developed for congestion quantification enables Royal FloraHolland to actively monitor the time they waste on waiting times in the future.

9.3 Identification

Classifying the origins of congestion through literature study and data analysis.

Five possible causes of congestion in the reversed order picking system of Royal FloraHolland were identified (page 39). Previous studies and high correlation coefficients with the number of potential pick face blockings indicated that the number of stops, total time spent on transferring products and the number of workers are likely to cause congestion. Moreover, the number of utilized locations indicates the size of the warehouse, which is a relevant factor for congestion. Lastly, managers at Royal FloraHolland were interested in the effect of new workers in the system. They were convinced that new workers may cause congestion as they are not familiar with the process.

9.4 Regression

Development of a tool that explains the amount of congestion in the system.

The number of potential pick face blockings per five-minute interval in the warehouse was used as dependent variable in regression analysis. All datasets from Week 35 were used in order to test the explanatory power of the identified causes. This resulted in a model that is able to explain 94% of the variability in the number of potential pick face blockings (page 47). With this tool, Royal FloraHolland is able to calculate an expected number of potential pick face blockings through total number of stops, locations and the time spent on transferring products.

Concluding, this research provided Royal FloraHolland with new methods and tools to retrieve and evaluate valuable information on congestion from their existing datasets. The regression modelling showed that three variables of the reversed order picking system can explain a significant part of the congestion. Moreover, the visualization methods pointed towards an uneven distribution of activities in the system. With respect to existing literature, this thesis created new added value in the form of potential pick face blocking identification and quantification of waste due to congestion. Also, the regression modelling showed that transfer time has more explanatory power on congestion than the number of workers in the system, which is identified as the foremost cause of congestion in existing literature.

9.5 Future Research

Further research into the three identified variables that cause congestion should indicate the possibilities that Royal FloraHolland has to reduce waiting times. Also, new research should consider tracking workers through an Indoor Positioning System (IPS) which would generate data that is less subject to workers that do not follow regular routines. Such an IPS can be put in place by Ultra Wide Band (UWB) technology (Silvia, Martina, Fabio, & Alessandro, 2018). Through UWB, the electric trucks can be traced and if this technology can be matched with RFH's Warehouse Management System, it can identify when and for how long pickers are queueing. By means of this, RFH will be able to measure all forms of congestion, instead of solely pick face blockings which was the case in this thesis' research. Regardless, this thesis created valuable insights in the congestion patterns in a reversed order picking system. Moreover, it showed that data mining is able to retrieve extremely important information from warehouse operations that had remained unobserved up until this research was conducted.

CHAPTER 10 DISCUSSION

This chapter discusses the results that were found by the research on Royal FloraHolland's reversed order picking system. Moreover, it discusses the limitations that come with these results. Based on these limitations, recommendations are made for future research.

First of all, the research conducted for this thesis resulted in exactly what was asked for by Royal FloraHolland. They have been provided with new methods that enable identification and quantification of congestion in their system. Moreover, the regression model has specified the directions for future research on the reduction of congestion.

This thesis proved that on Monday the 26th of August, 18% of all stops that were made within the warehouse experienced potential pick face blocking. In total, this resulted in nearly 30 manhours of waste due to congestion. These pick face blockings were caused by the number of stops, total transfer time and the number of locations that were utilized. Surprisingly, the warehouse is characterized by highly congested locations and locations with only a few transactions throughout the whole process. This points out that the clustering of activity has major effects on congestion. Therefore, it is recommended that both Royal FloraHolland and future academic literature aim at determining the precise effect of this clustering on congestion.

Unfortunately, this thesis was limited to the logged data that was available. Even though extensive data mining resulted in key insights, identifying pick face blockings that occur due to queueing lines was not possible. Managers pointed out that this form of congestion also plays a role in the distribution process. However, the identification of such congestion would have become too much of a derived variable with many insecurities. Also, the potential pick face blocking variable developed in this thesis was not able to recognize this congestion. The best way to analyze this type of congestion is through Ultra Wide Band technology. By means of this, workers that are waiting to access a location which is blocked by a queueing line can be flagged. Concluding, this thesis leaves a great opportunity for future research to design methods that are able to identify pick face blockings caused by queuing lines.

CHAPTER 11 EPILOGUE

This chapter gives a brief overview of what I learned during my internship and shows the importance of my findings for the daily operations of Royal FloraHolland.

The research that I conducted at RFH challenged my capabilities, especially in the field of data gathering. As my research progressed, it turned out that I needed more data than was expected beforehand. Gathering this data was something completely new for me and my thesis supervisor (Mr. Dekker) encouraged me to master R, such that I could retrieve the information I was looking for from the logfiles of RFH's voice system.

The additional timestamp that I retrieved from the data allowed me to set up the methods for the identification of potential pick face blocking and the quantification of related congestion. Additionally, the model that was used to explain congestion in the reversed order picking system of RFH was based on the newly gathered data.

As the managers at RFH were impressed with the possibilities that came with this additional timestamp, they decided to write up a request to the IT-department. When I completed my internship at RFH, the IT-department was working very hard to store the new timestamp in the daily reports. With these reports, the managers at RFH will be able to identify and quantify congestion in the system on a daily basis.

Since the managers at other facilities of RFH also showed curiosity regarding my research, I presented my results to them as well. Their enthusiasm was overwhelming and they promised me that they were going to implement the methods that I created during my internship. All in all, I feel very honoured with the great interest in my research. In the end, this underlines the relevance and success of my research.

BIBLIOGRAPHY

Applegate, D. L., Bixby, R. E., Chvatal, V., & Cook, W. J. (2011). *The traveling salesman problem : A computational study.* Princeton: Princeton University Press. doi:10.1515/9781400841103

Bartholdi, J. J., & Hackman, S. T. (2008). *Warehouse & distribution science : Release 0.89* (Rev. [ed.]. ed.). Atlanta, Ga.: The Supply Chain and Logistics Institute.

Beamon, B. M. (1999). System reliability and congestion in a material handling system. *Computers & Industrial Engineering, 36*(3), 673-684. doi:10.1016/S0360-8352(99)00158-8

De Koster, R., Le-Duc, T., & Roodbergen, K. J. (2007). Design and control of warehouse order picking: A literature review. *European Journal of Operational Research, 182*(2), 481-501.

de Vries, J., de Koster, R., & Stam, D. (2016). Exploring the role of picker personality in predicting picking performance with pick by voice, pick to light and RF-terminal picking. *International Journal of Production Research, 54*(8), 2260-2274. doi:10.1080/00207543.2015.1064184

Goetschalckx, M., & Ashayeri, J. (1989). Classification and design of order picking. *Logistics World, 2*(2), 99-106. doi:10.1108/eb007469

Grosse, E. H., & Glock, C. H. (2015). The effect of worker learning on manual order picking processes. *International Journal of Production Economics, 170*(Part C), 882-890.

Hand, D. J. (1998). Data mining: Statistics and more? *The American Statistician, 52*(2), 112-118.

Hand, D. J., Blunt, G., Kelly, M. G., & Adams, N. M. (2000). Data mining for fun and profit. *Statist.Sci., 15*(2), 111-131. doi:10.1214/ss/1009212753

Hsieh, L., & Tsai, L. (2006). The optimum design of a warehouse system on order picking efficiency. *International Journal, Advanced Manufacturing Technology, 28*(5), 626-637.

Huber, C. (2014). *Throughput analysis of manual order picking systems with congestion consideration.* PhD thesis, Karlsruher Institut für Technologie.

Le-Duc, T. (. (2005). *Design and control of efficient order picking processes.* Thesis dissertation, Erasmus University Rotterdam.

Parikh, P. J. (2006). *Designing order picking systems for distribution centers.* (Ph.D. thesis, Virginia Tech).

Parikh, P. J., & Meller, R. D. (2009). Estimating picker blocking in wide-aisle order picking systems. *IIE Transactions, 41*(3), 232-246.

Sandbrink, J. (2016). *Worker congestion in a picker-to-parts warehouse: A case study.* Thesis dissertation, TU Eindhoven.

Schrotenboer, A. H., Wruck, S., Roodbergen, K. J., Veenstra, M., & Dijkstra, A. (2017). Order picker routing with product returns and interaction delays. *International Journal of Production Research, 55*(21), 6394-6406.

Silvia, Z., Martina, C., Fabio, S., & Alessandro, P. (2018). Ultra wide band indoor positioning system: Analysis and testing of an IPS technology. *IFAC PapersOnLine, 51*(11), 1488-1492. doi:10.1016/j.ifacol.2018.08.292

Tompkins, J. A. (2010). *Facilities planning* (4th ed. ed.). Hoboken, NJ; Chichester: John Wiley & Sons ;; John Wiley [distributor].

van Harn, K., & Steutel, F. W. (1995). Infinite divisibility and the waiting-time paradox. *Stochastic Models*, *11*(3), 527-540.

Wooldridge, J. M. (2016). *Introductory econometrics : A modern approach* (Sixth edition, student edition. ed.). Boston, MA: Cengage Learning.

APPENDICES

Appendix I – Pick Face Blocking Parameter

Within this thesis, it is assumed that workers interacted at the same warehouse location if they arrive at most fifteen seconds after their predecessor has left. This limit for the time difference between two workers was set after the congestion quantification took place. Congestion indicates that a worker has exceeded theoretical travel time for a given distance. When examining congestion, it becomes clearly visible that congestion occurs much more for low values of time difference. This is supported by Figure A-1, which is a subset of a random 10-minute interval from August 26.

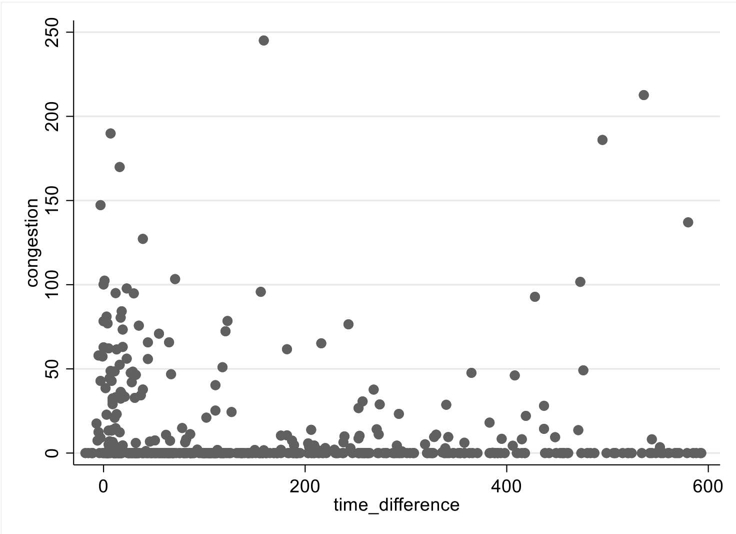

Figure A-1 – Scatterplot congestion over values for time difference

Also, when separating all data from August 26 into subsets based on their values for time difference, interaction seems to take place until the boundary of fifteen seconds is reached (Table A-1).

Table A-1 – Congestion per observation by time difference on August 26

Time difference (t)	$t < 5$	$5 < t < 10$	$10 < t < 15$	$15 < t < 20$	$20 < t < 25$
Average congestion per observation (in seconds)	24.3	22.8	21.5	16.3	12.5

Since the congestion shows a relatively large drop after fifteen seconds, this was set as the limit for the identification of pick face blocking.

Appendix II – OLS-assumptions Theoretical Travel Times

This Appendix discusses the Ordinary Least Squares assumptions with respect to the model of theoretical travel times.

Linear in Parameters

There is no reason to believe that travel time is a non-linear combination of travel distance and the error terms in the model.

Random Sampling

It is quite possible that the first ten minutes of the distribution process have different characteristics compared to the rest of the process. For example, it has been determined that not all electric trucks have the same acceleration and top speed. It might be the case that mostly fast trucks are used in the first part of the process. If workers know which trucks are fast and which ones are slow, there might be a problem with the random sampling assumption. Also, the process starts with only permanent employees and no temporary workers. This could indicate that the theoretical travel times established in this model have a downward bias, since permanent workers are able to work fast and with fewer delays than temporary workers.

To overcome this problem, the same regression procedure was followed for a random ten-minute interval on the 26th of August. Following exactly the same steps resulted in nearly the same coefficients for the regression model. Thus, indicating there was no violation of the random sampling assumption.

No Perfect Collinearity

Since there is only one explanatory variable included in the model for theoretical travel times, there is no reason to investigate whether this assumption is violated.

Zero Conditional Mean

To test whether travel time can be explained by a variable that is also correlated with travel distance, the number of turns within each trip was incorporated into the model. Unfortunately, this did have significant explanatory power. Also, the meters that were driven on the cross aisle compared to the meters driven within aisles did not have any effect. This was based on the assumption that trips with relatively more meters on the cross aisle and only a few turns may reach a higher average speed. Therefore, it does not seem as if this assumption is violated.

Homoskedasticity

Logically, after creating the band of 20 seconds in the model all observations are randomly scattered around the regression line. Thus, the homoskedasticity assumption is not violated.

Normally Distributed Error Terms

When plotting the residuals from the model against their fitted values, it becomes apparent that there is no clear concentration of observations along the regression line. As shown in Figure A-2, there are almost exactly the same number of observations underneath as above the regression line.

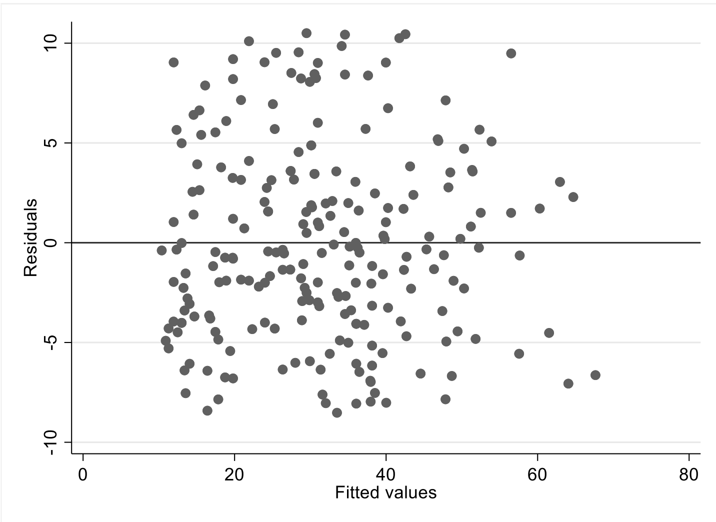

Figure A-2 – Residuals theoretical travel times model

This non-normal distribution is likely to be the result of the manipulation of the dataset. When deleting observations above the threshold, this is what is left over. However, there was never expected to be a normally distributed error term. This is due to the different characteristics of electric trucks and the occurrence of traffic for some observations whilst others had no traffic at all. As these causes might unfairly create an upward bias in travel time for all observations when fitting a regression line, it seems better to adjust the line to the upper bound of the data in Figure A-2. By means of this, congestion is only quantified if observations exceed the threshold of the fastest observation plus 20 seconds. In the end, this does not make a difference for congestion since this is classified as the difference between actual travel times of observations with and without pick face blocking.

Appendix III – OLS-assumptions Congestion Predictions

This Appendix discusses the Ordinary Least Squares assumptions with respect to the models that are discussed in Chapter 8. The first three assumptions that are discussed, consider all three models. Thereafter, each assumption is tested per model. The methods that are used for model evaluations are based on theory from Wooldridge (2016).

Linear in Parameters

All models discussed are linear in their parameters. Model II includes a variable that contains a nonlinear relationship, however this does not influence the linearity in parameters. The number of pick face blockings was always estimated as a linear function of the explanatory variables.

Random Sampling

Each model uses the same random set of data. The aggregated observations consider Week 35 of 2019, which does not seem to be non-random at all. The only difference with weeks in the peak-season is that the process might have lasted a little bit shorter. However, there is no reason to believe that the number of pick face blockings reacts differently to changes in the explanatory variables compared to other weeks.

No Perfect Collinearity

None of the variables included in the models is a perfect linear function of another variable. Thus, this assumption is not violated. Also, if there would have been perfect collinearity, the Stata package would have omitted one of the variables causing perfect collinearity.

Model I

Zero Conditional Mean

With respect to this model, a plot showing the residuals compared to the fitted values indicates that the mean of the error term is not zero for each value of the number of stops (Figure A-3).

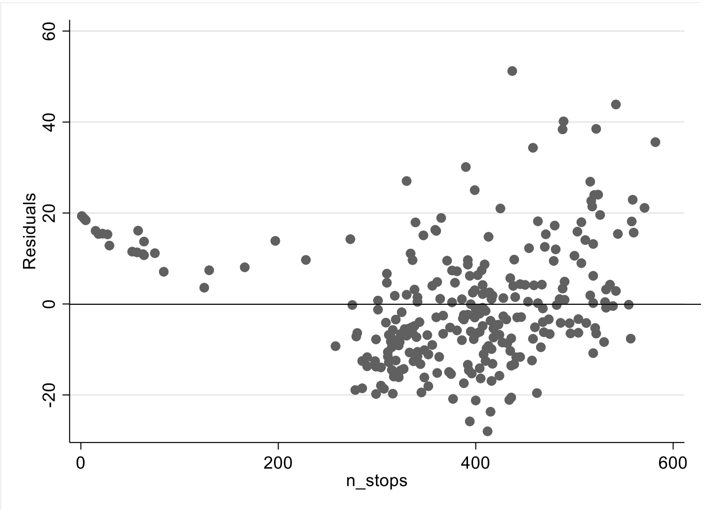

Figure A-3 – Zero Conditional Mean violation model I

Obviously, the mean of the residuals shows a value higher than zero for the lowest and highest values of the number of stops. Next to this, the mean of the residuals is below zero in the middle of the plot. Thus, the Zero Conditional Mean (ZCM) assumption is violated for model I.

Homoskedasticity

From Figure A-3 it can be deducted that the variance of the residuals becomes larger for larger values of the number of stops, indicating heteroskedasticity. Also, White's test on homoskedasticity (Wooldridge, 2016) gives a p-value of 0.001, pointing towards heteroskedasticity.

Normally Distributed Error Terms

The error terms from model I show a right-skewed pattern, as shown in Figure A-4. The

Kurtosis test for normality (Wooldridge, 2016) supports this finding. The p-value of 0.000 from the Kurtosis test indicates that the residuals do not fit a normal distribution.

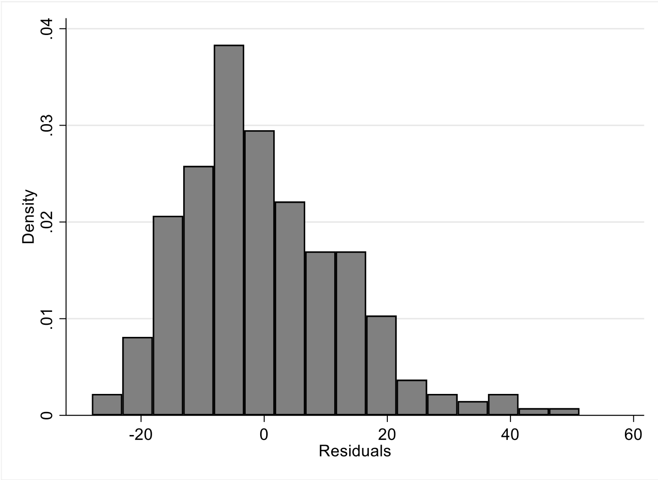

Figure A-4 – Histogram error terms model I

Concluding, three of the six OLS-assumptions are violated by model I. Therefore, this is not a good model to explain the number of pick face blockings in the process.

Model II

Zero Conditional Mean

Due to the inclusion of the quadratic term, the violation of the ZCM-assumption by model I is solved (Figure A-5). All residuals are scattered around a mean of zero, indicating no problems regarding this assumption. However, it could still be the case that there is a form of Omitted Variable Bias.

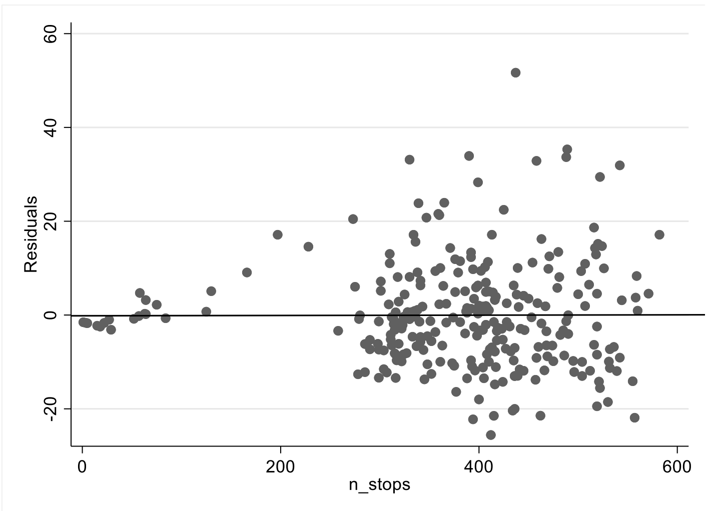

Figure A-5 – No Zero Conditional Mean violation model II

Homoskedasticity

Similar to model I, the second model shows heteroskedastic error terms. The scatterplot shows a funnel shape (Figure A-5). Also, White's test generates a p-value of 0.042, indicating that the residuals have improved but still show heteroskedasticity.

Normally Distributed Error Terms

Figure A-6 shows that the residuals of model II are still right-skewed. Again, this is supported by a p-value of 0.000 for the Kurtosis test on normality. This makes the second model just as unsuitable for prediction of the number of pick face blocking as model I, even though the problem of ZCM-violation is solved.

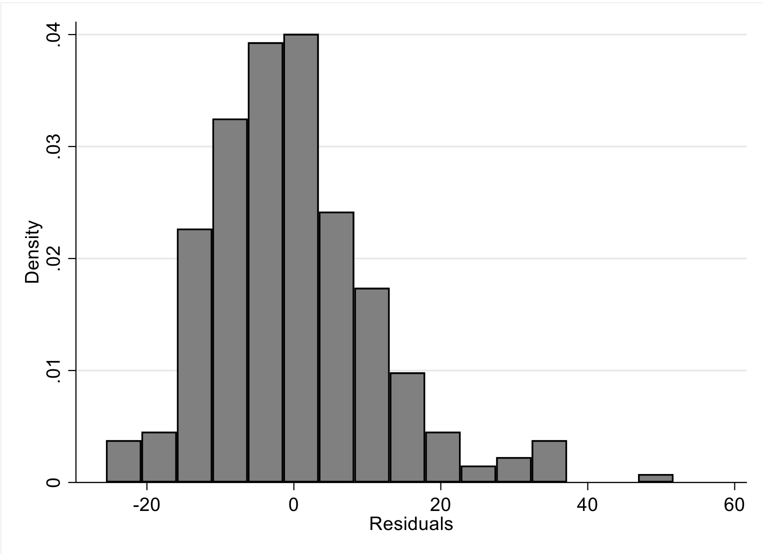

Figure A-6 – Histogram error terms model II

Model III

Zero Conditional Mean

The three variables included in the third model induce a slight violation of the ZCM-assumption at the lowest values of the number of stops in the process (Figure A-7). Besides these observations, the error terms return a mean of zero.

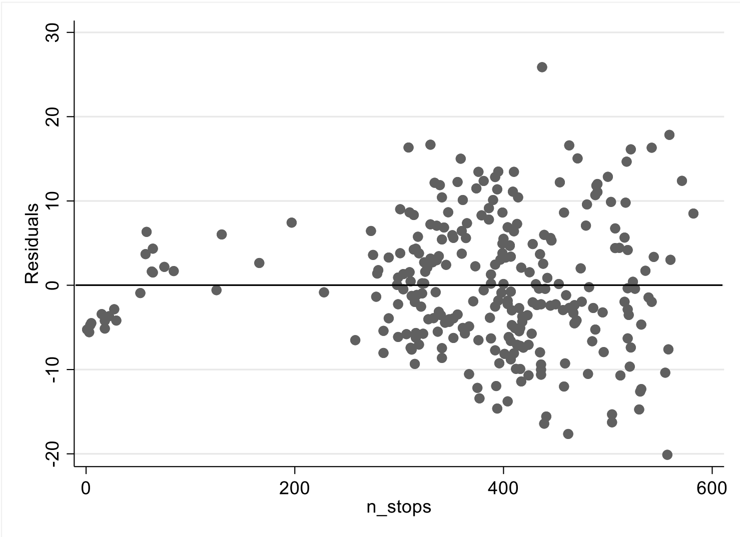

Figure A-7 – No Zero Conditional Mean violation model III

Homoskedasticity

Figure A-7 also shows that the funnel-shaped pattern of the residuals remains apparent. Again, a p-value of 0.000 from White's test indicates there is heteroskedasticity in the error terms. Therefore, White's robust option (Wooldridge, 2016) was used for the output of the final model (as this model will be used for future reports at RFH) to account for the heteroskedasticity in the error terms.

Normally Distributed Error Terms

After using White's robust option, the residuals from model III follow a normal distribution (Figure A-8).

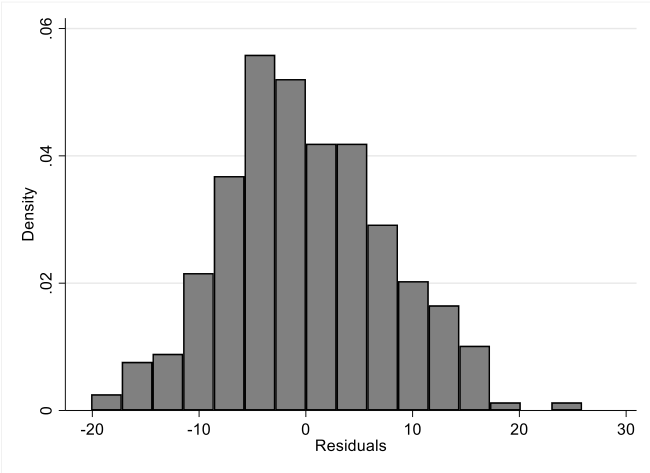

Figure A-8 – Histogram error terms model III

The skewness from the previous models is solved and the Kurtosis test on normality returns a p-value of 0.285, indicating that de residuals approximately follow a normal distribution.

Concluding, after correcting for heteroskedasticity in the error terms by using White's robust option, the third model violates none of the OLS-assumptions. The residuals even tend to follow a normal distribution which leads to a proper inference of the model's parameters.