Eindhoven University of Technology

MASTER

The effects of decisional guidance on the forecast performance in judgmentally adjusting forecasts

Geurts, Jochem M.

Award date:
2022

Eindhoven University of Technology
Department of Industrial Engineering & Innovation Sciences

Master thesis

The effects of decisional guidance on the forecast performance in judgmentally adjusting forecasts

Full Name	Student ID	Study
J.M. Geurts	1511785	Operations Management & Logistics

Supervisors:
dr. C. Imdahl
dr. ir. R.J.I. Basten
Company supervisors:
dr.ir. B. van der Staak
E. Versteijnen

Eindhoven, November 25, 2022

This page is intentionally left blank.

Abstract

In this thesis, we evaluate the impact of forecast adjustments on performance as literature indicates that planners do not always add value to the forecast accuracy. We test whether decisional guidance can lead to improved forecasting performance. To test where planners add value, we first evaluated the performance. To this end, we did perform data analysis on judgmentally adjusted forecasts. Results show that, on average, the judgmental adjustments do not improve the accuracy of the statistical forecasts. This is partly due to the excellent performance of the statistical forecasts and the presence of cognitive biases. Overall, planners are good at choosing the right direction but have difficulties with selecting the right size of an adjustment. Furthermore, planners improve the forecast accuracy for items with high volatility but do not add value for items with medium to low volatility. To improve the performance of the judgmental adjustments, decisional guidance is tested in an experiment to help the planners with their judgments. The findings of the data analysis were implemented in the experiment. We conclude that using the decisional guidance results in a significantly lower forecast error and can improve the current situation.

Preface

After more than six years of studying in Tilburg, Eindhoven, Santiago and Valencia, my study time is ending. After studying International Business Administration in Tilburg, I made a choice to extend my knowledge about the operations and technical processes within companies and study Operations Management and Logistics. My interest in optimizing processes, combined with the knowledge obtained during my bachelor's, is combined in the subject of this master thesis. Therefore, this master thesis project felt like a good end to my time as a student.

I would like to start by thanking my supervisors from the University. I want to thank Christina Imdahl for guiding me through this process over the past months. I value that she provided me with all the feedback and all her knowledge during this master thesis. Secondly, I would like to thank Rob Basten for finding a suitable company for me. Furthermore, I want to thank you for your provided feedback on my thesis.

I also would like to thank Bregje van der Staak and Edward Versteijnen from EyeOn. I want to thank Bregje for being the perfect link between University and EyeOn. Furthermore, I would like to thank Edward for providing me with all the business insights and his feedback. At last, I want to thank Maarten van Liempd for answering all my questions.

Finally, I would like to thank my friends and family for their support during my entire study and this final project. I want to thank all my friends who made my student life in Tilburg an unforgettable time. Last, I want to thank my friends who supported me during my final project. Thanks for the advice and support during the time we were all writing our master thesis!

Management Summary

Introduction

Over the last few years, the literature on judgmentally adjusting forecasts has been increasing. There is a consensus that judgmental adjustments by planners add value to forecast accuracy. However, there is also proof that judgmental adjustments are not perfect, and often unnecessary adjustments are performed (Lawrence et al., 2006). As a possible way of improving the process of judgmentally adjusting forecasts, research on the effects of providing feedback or decisional guidance is often mentioned as an interesting direction in literature (Petropoulos et al., 2017; Fildes et al., 2006). To get a better understanding of the exact performance and behavior of planners, the first research question is:

> **Research Question 1:** What kind of behavior of planners in judgmental forecasting should be guided by decisional guidance?

There are multiple forms of decisional guidance mentioned in the literature, such as suggestive and informative guidance (Montazemi et al., 1996). Suggestive guidance implies proposing action to the decision maker, while informative guidance gives unbiased, relevant information to the decision maker without any suggestions on actions to take. In this thesis, the effect of these forms has been tested. Therefore, the second research question is:

> **Research Question 2:** What form of guidance is most effective for improving the performance of judgmentally adjusting forecasts?

The objective of this thesis is to gain insights into the current performance and behavior of planners and to test how decisional guidance can add value to forecast performance. The thesis is performed at EyeOn.

Literature review

A literature review is conducted to provide background information of the theoretical foundation to the reader about judgmentally adjusting forecasts and decisional guidance. Literature has focused on forecasting for many years. For companies, forecasting sales is about making accurate predictions of the actual sales of certain products over a certain period (Ritzman & King, 1993). The predictions of the sales influence many other decisions in many aspects of the supply chain, such as inventory management (Fildes et al., 2009). This highlights the importance of an accurate forecast.

First, we give an overview of the forecasting process. Companies make their forecasts in multiple ways. It is done manually by a judgmental forecast of the demand planners, based on a statistical forecast or as a combination of the two previously named methods (Nahmias & Olsen, 2015). Many companies use the combination of statistical forecasts with judgmental adjustments. The process of judgmentally adjusting the forecast consists of two phases (Arvan et al., 2019). First, the planner needs to decide whether an adjustment of the forecast generated by the system is necessary. When the planner has decided an adjustment is necessary, the second phase consists of deciding on the direction and the size of the adjustment.

To provide an accurate forecast, it is important that a planner has background information about the items. In order to provide a planner with information on the importance and predictability of products, often a classification of items is used (Scholz-Reiter et al., 2012). Classification of items can be done in multiple ways, but in many cases, the classification uses the ABC and XYZ analyses. The ABC analysis ranks the products based on the annual turnover, and the XYZ analysis ranks products based on the level of uncertainty of demand (Scholz-Reiter et al., 2012).

The adjustments of statistical forecasts by planners are conducted to improve the forecast accuracy (Li & Jiang, 2017). Many articles show substantial evidence that including a judgmental adjustment focusing on special events not taken into account by the statistical forecast can improve the forecast accuracy (Fildes et al., 2009). The current issues and dynamics in the current business settings make it very difficult to only rely on statistical forecasting methods (Alvarado-Valencia et al., 2017). However, there is also a significant amount of literature on the characteristics that could negatively affect the performance of a planner in adjusting the forecast. One of the main disadvantages of judgmental forecasting described in many studies is that planners can be biased when making adjustments (Sanders & Ritzman, 2004).

Cognitive biases in judgmental forecasting can cause lower performance when adjusting statistical forecasts (Fildes et al., 2009). The research of Eroglu & Croxton (2010) focuses on the optimism bias, the anchoring bias and the overreacting bias. The optimism bias refers to the tendency to project mainly positive sales results in the future, which leads to mainly positive errors (Eroglu & Croxton, 2010). The anchoring bias implies the situation of a forecaster adjusting statistical forecasts in the right direction, but the adjusted value stays too close to the statistical forecast as an anchor value (Eroglu & Croxton, 2010). Finally, the overreaction bias implies when a planner adjusts in the right direction but overshoots the actual sales resulting in a too large adjustment that increases the error (Eroglu & Croxton, 2010).

When people make decisions, decisional guidance can be provided by a decision support system (DSS) to influence or support the users in their decision-making process. Decisional guidance is how a DSS enlightens the decision-makers with structuring and executing the decision-making processes (Silver, 1991). Decisional guidance helps to structure the decision-making task and to execute the task (Silver, 1991). A DSS can provide decisional guidance on purpose, deliberate guidance, or when unintentionally, inadvertent guidance (Silver, 1991). Since deliberate guidance intends to influence the decision-maker, this is the main topic studied in earlier literature. Deliberate guidance can be divided into suggestive and informative guidance (Montazemi et al., 1996; Fildes et al., 2006). Suggestive guidance implies proposing action to the decision maker, while informative guidance gives unbiased, relevant information to the decision maker without any suggestions on actions to take. The research of Montazemi et al. (1996) found that suggestive guidance outperformed informative guidance when completing a less complex task. With a more complex task, participants receiving informative guidance outperformed those receiving suggestive guidance according to the study of Montazemi et al. (1996).

Data analysis

To answer the first research question, a dataset of judgmental adjustments from a customer of EyeOn has been used. After cleaning and filtering steps, an outlier check was performed. The data analysis showed that planners, on average, do not add value to the forecast with their adjustments. Planners are quite good at determining the right direction of an adjustment, but this does not always lead to improved forecast accuracy due to overreacting. This means that planners make an adjustment that is too large compared to the actual sales. Furthermore, when improving the forecast accuracy, the added value of the adjustment is not very high due to anchoring. This means that planners adjust in the right direction but stay too close to the statistical forecast. Results furthermore showed the difference in performance among XYZ categories. Planners, on average, improved the forecast accuracy for Z-items but did not improve the forecast accuracy for X and Y-items.

Experiment

With the results of the data analysis, an experiment has been set up to test the effects of decisional guidance on forecast performance. In the experiment, participants got the task to review a statistical forecast with additional, more accurate information compared to the statistical forecast and, if deemed necessary, adjust the forecast. This was tested on forecasting experts of EyeOn. In total, 36 people participated. They received multiple forms of decisional guidance to test the effect of the different forms of guidance on forecast accuracy.

Results experiment and recommendation

The results of the experiment showed that participants were willing to use the decisional guidance and improved forecast accuracy because of the decisional guidance they received. For Y and Z-items, the decisional guidance on the size of the adjustment had a larger effect compared to X-items in terms of improving the forecast accuracy. Suggestive guidance on the adjustment had the most impact on X-items in terms of added value to the accuracy. For those items, there were no significant differences found compared to the guidance on size. Furthermore, planners were more inclined to accept the suggestive guidance on the adjustments for X-items compared to Y and Z-items. In general, the results show positive effects of decisional guidance on forecast accuracy. Therefore, it is suggested to use decisional guidance to improve forecast accuracy. The results furthermore showed that not all provided guidance leads to an improved forecast due to the quality of the guidance and the willingness of participants to accept the guidance. Therefore, future research on the computational side of decisional guidance and more specific research on when exactly to use the guidance could be very helpful in improving forecast accuracy.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Company description

This master thesis is conducted at the Dutch consulting company EyeOn in Aarle-Rixtel. EyeOn is a forecasting and planning consultancy company with more than 20 years of experience, which gives them a lot of knowledge about improving and implementing forecasting and planning processes and systems. EyeOn provides insights to customers about their planning processes. They guide companies in consultancy projects and forecasting and planning projects as a service. Currently, the main focus industries of EyeOn are Life Science, Process, Consumer Products and Complex Products & Systems (CoPS). At the moment, EyeOn has around 90 employees in 6 different locations worldwide. EyeOn has four industry teams that are active in the four industries just mentioned. Furthermore, EyeOn has 3 product teams, Data Science, Solutions and Planning Services. This thesis project is part of the product team Planning Services. Planning Services is a team that "provides robust, recurring outsourced services focused on high-quality forecasting, inventory optimization and actionable end-to-end supply chain insights" (EyeOn, n.d.).

1.2 Judgmentally adjusting forecasts

In every supply chain, it is essential and valuable to generate accurate forecasts of the demand to make important decisions and schedule according to the demand (Lawrence et al., 2006). Accurate forecasts have consequences for all levels in a supply chain (Fildes et al., 2009). Inaccurate forecasts result in poor service levels or excess inventory, leading to high costs (Fildes et al., 2009). The forecasting task is difficult due to the existence of outliers, level and trend shifts, and the impact of the market and the economic environment (Fildes et al., 2009). In general, forecasting demand involves the use of a system-generated forecast. When statistical models do not capture all dynamics of the business context because of information about, for example, promotions or new product launches at a competitor incorporated in the system, human judgment is a common approach to add value to the forecast (Arvan et al., 2019). Using human judgment in forecasting can be combined with statistically generated forecasts. A popular method to include this human judgment is by adding the possibility of judgmental adjustments after the system has generated the forecast. When a system generates forecasts, planners can look at those forecasts and decide whether they want to adjust them. They can make an upward or downward adjustment if they want to adjust. An illustration of judgmental adjustments in forecasting is visible in Figure 1.1. Through this way of working, planners can use their specific expertise and knowledge about events and changes not incorporated in the system to improve the accuracy (Fildes et al., 2009). Research and case studies done in the past prove the value that those judgmental adjustments can add to the forecast accuracy (Fildes et al., 2009; Franses & Legerstee, 2009; Arvan et al., 2019).

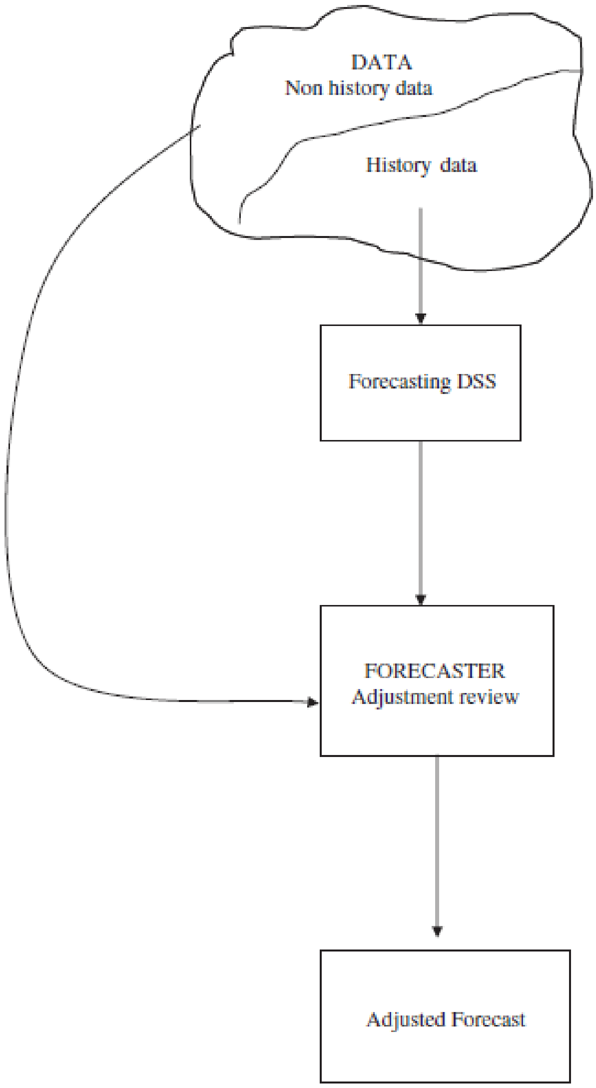

Figure 1.1: Judgmental Forecast (Lawrence et al., 2006)

The importance of forecasting is evident in the entire supply chain. Forecasting accurate numbers of expected demand influences many other decisions in many aspects of the supply chain, such as inventory management. Therefore, there often is a collaboration between forecasting and inventory management in a supply chain or company. For forecasting, it is important to have background information about inventory management. One of the components of the information is the importance of a particular demand item. In supply chain inventory management, it is common to classify the demand items to indicate the importance of demand (Scholz-Reiter et al., 2012). Classifying items can be done in multiple ways, but in many cases, the classification uses the ABC and XYZ analysis. The ABC analysis usually ranks the products based on the annual turnover, and the

XYZ analysis ranks products based on the level of uncertainty of demand (Scholz-Reiter et al., 2012). The XYZ analysis ranks products based on how difficult it is to predict the demand for a certain item. This classification is often used in judgmental forecasting to understand better which items are most important for planners to adjust judgmentally.

1.3 Problem

Besides the positive effects of judgmental adjustments, there is also proof that the judgmental adjustments are not perfect and planners often make unnecessary adjustments (Lawrence et al., 2006). Fildes et al. (2009) found in their research that judgmental forecasts are biased and inefficient but improve forecast accuracy. Furthermore, their study showed that especially minor adjustments should be avoided. Articles of Fildes et al. (2007), Franses & Legerstee (2009) and Fildes et al. (2009) demonstrated the effects biases have in judgmental adjusting forecasts. A bias is a systematic deviation from some standard (Eroglu & Croxton, 2010). The efficacy of a judgmental adjustment to a forecast depends on multiple things, of which the willingness to integrate information into the forecast is a vital aspect (Eroglu & Sanders, 2021). As stated above, biases can occur when planners judgmentally adjust forecasts. Research of Eroglu & Croxton (2010) concluded the significant effect personality of a forecaster could have on the biases. The phenomenon of different forecasting behavior among individuals is forecasting heterogeneity (Pennings, 2016). Research of Pennings (2016) also concluded that individual biases could substantially influence the accuracy of judgmental adjustment. However, biases in judgmental forecasting often have been studied based on aggregate results (Schweitzer & Cachon, 2000; Bolton & Katok, 2008). The studies based on aggregate results imply the data of all the forecasters are combined. Since individual biases can influence the forecast substantially, the research focus on aggregate results can cause problems. A method used in literature to distinguish between individuals in judgmental forecasting is the usage of biases in the research of Eroglu & Croxton (2010). The research of Eroglu & Croxton (2010) focuses on the optimism bias, the anchoring bias and the overreacting bias. The optimism bias refers to the tendency to project mainly positive sales results in the future, which leads to particularly positive errors Eroglu & Croxton (2010). The anchoring bias implies the situation of a forecaster adjusting statistical forecasts in the right direction, but the adjusted value stays too close to the statistical forecast as an anchoring value Eroglu & Croxton (2010). The overreaction bias implies when a planner adjusts in the right direction but overshoots the actual sales resulting in a too large adjustment that increases the forecast error (Eroglu & Croxton, 2010).

There is much to improve in the current adjustment process. Fildes et al. (2006) found that effective design features in forecast support systems can improve forecast accuracy. On the topic of decisional guidance, Petropoulos et al. (2017) concluded in their research that providing feedback could be a very effective method to improve forecast accuracy, and it could be valuable to include such an element in the design of the forecast support system. Petropoulos et al. (2017) recommend investing time and effort to research the topic of decisional guidance and training of planners that make the adjustments. Decisional guidance implies the influencing and support offered during a decision-making process often done by the decision support system (Parikh et al., 2001). Multiple articles studied the effects that decisional guidance can have on a decision-making process (Parikh et al., 2001; Silver, 1991; Montazemi et al., 1996; Fildes et al., 2006). Decisional guidance can provide information during the decision-making process that can help the user in structuring and executing the decision-making task (Silver, 1991). Since it is recommended in many papers (Petropoulos et al., 2017; Lawrence et al., 2006; Fildes et al., 2009) to further investigate the effect feedback and guidance can have on the accuracy of judgmental adjustments in combination with existing literature on decisional guidance (Parikh et al., 2001; Silver, 1991; Montazemi et al., 1996; Fildes et al., 2006), decisional guidance on judgmentally adjusting forecasts will be the main subject of this thesis.

1.4 Research questions

As described in the previous section, many reviews stated the added value of judgmentally adjusting forecasts. Furthermore, many adjustments have an undesired outcome regarding forecast accuracy and bias. Multiple reviews suggest the research direction of feedback and decisional guidance to improve the process of judgmentally adjusting forecasts. Therefore, the research question of this master thesis is:

> **Research Question:** How can decisional guidance be implemented in decision support systems to improve the performance of judgmental adjustments in forecasting?

To gain more insights in the current performance of the judgmental adjustments overall, in the different segmentation categories and among the different type of planners the first part contains the following questions:

> **Research Question 1:** What kind of behavior of planners in judgmental forecasting should be guided by decisional guidance?

The focus in this first part will be on the behavior of planners specifically among items from different segmentation categories ranked by the ABC/XYZ analysis. Furthermore, specific focus is on the differences of planners, and the biases they have. Therefore, the more detailed questions are the following:

> *Research Question 1a:* How does the performance of planners in judgmental forecasting differ among segmentation categories?

> *Research Question 1b:* How does the performance of planners in judgmental forecasting differ among planners?

With the knowledge about the current performance, in the second part the decisional guidance component will be tested and analyzed resulting in the following research question:

> **Research Question 2:** What form of guidance is most effective for improving the performance of judgmentally adjusting forecasts?

This part will focus on conducting an experiment in order to find out which form of decisional guidance increases the accuracy of judgmental forecasting the most. This will be analyzed by overall performance, as well as more specific details regarding performance in different segmentation categories and among different planner types. Therefore, the following question is defined to provide more guidance during this thesis:

> *Research Question 2a:* What form of decisional guidance is most effective for improving the performance of judgmentally adjusting forecasts among the different segmentation categories?

1.5 Research setting

This master thesis is conducted in combination with the company EyeOn. EyeOn Planning Services delivers statistical forecasts to its customers on a regular basis. At the customer, demand planners can adjust the statistical forecast based on the information they have. The process at EyeOn is in line with the description of the forecasting process explained above. The only difference with the earlier provided description is that companies outsource the statistical forecasts to EyeOn instead of generating the statistical forecasts internally. For this specific case, Company A is the customer of EyeOn used for this thesis. Company A delivers its data sets to EyeOn. Afterward, EyeOn will review and correct the delivered historical data of Company A. Next, EyeOn conducts several steps such as outlier cleansing, seasonality determination, trend detection and portfolio segmentation based on the ABC/XYZ ranking. An example of the different product categories based on segmentation and the forecasting strategy of EyeOn is visible in Figure 1.2. Figure 1.2 shows which items planners should focus on and which items they should not focus. The thresholds used for the segmentation are visible in Tables 1.1 and 1.2. The numbers in Table 1.1 refer to annual turnover, and the numbers in Table 1.2 refer to the level of uncertainty of demand given by the coefficient of variation.

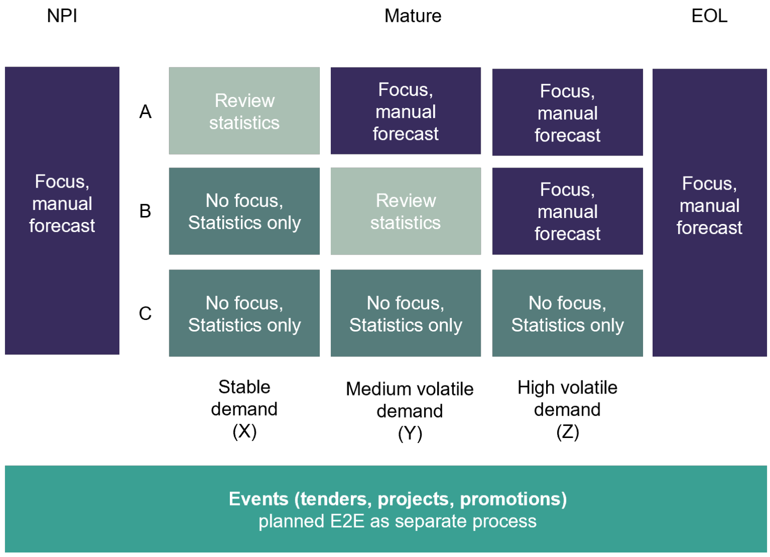

Figure 1.2: Segmentation categorization EyeOn

Table 1.1: Thresholds ABC segmentation

	Threshold
A-B	0.8
B-C	0.95

Table 1.2: Thresholds XYZ segmentation

	Threshold
X-Y	0.4
Y-Z	1

After these steps, EyeOn generates statistical forecasts. EyeOn generates statistical forecasts on Demand Forecast Unit (DFU) level. When the customer receives the statistical forecasts, demand planners can review them and decide to adjust them. At Company A, this happens in Jedox, a planning and performance management platform. Planners can make adjustments on the regional level, (sub)segment level, product family, product category, brand level, production line, banner and market/country level. The different levels are called hierarchy levels. Each business unit has its own hierarchy. This means that for one business unit, the regional level is hierarchy level 1, and for another, the regional level is hierarchy level 4. Planners can adjust the forecasts on multiple levels.

When reviewing the statistical forecasts and deciding if an adjustment is necessary, planners can see historical information about the demand, the statistical forecast, and the forecast before. The forecast before is the value that will be the final forecast if the planner decides not to adjust. In general, the forecast before will equal the statistical forecast. However, planners can adjust one item multiple times in one period. If the forecast before differs from the statistical forecast, someone adjusted that item earlier in that period. This earlier adjustment can be on a different hierarchy level but also on the same hierarchy level. When the planner decides to adjust, this value is the forecast after forecast after.

When making the adjustments, planners can add a comment to the adjustment overview such that the reasoning behind the adjustment is known. The addition of comments is an optional part of the adjustment process. In the current judgmental adjustment process, alerts warn the planners of possible significant errors. There are currently three types of alerts: the actual gap, the statistical gap and the zero gap. The actual gap shows the difference between the latest actual value, so the previous month, and the previous final forecast for the next month. The statistical gap alert identifies the items with the biggest difference between the statistical forecast of the current cycle with the final forecast from the previous cycle. The last type of alert is the zero gap alert. The zero gap alert identifies items with a previous positive forecast but no actual sales in the previous month. However, it is unknown whether the alerts are often used and help the planners in making adjustments. Therefore, the use of decisional guidance can be a way to improve this process of judgmentally adjusting forecasts.

To be able to answer research question 1, data analysis is performed on a dataset containing adjustments made by planners of Company A. The performance of the adjustments is analyzed among different situations to get a good overview of the current planners' performance in judgmentally adjusting forecasts. To be able to answer the second research question, an experiment is conducted to test the effects of decisional guidance on the performance when judgmentally adjusting forecasts. The goal is to get insights into the effects of the guidance in different situations and for different product characteristics.

Chapter 2

Literature review

In this chapter, literature on the topics relevant to this master thesis is reviewed and explained. First, forecasting literature is discussed with the main focus on judgmental forecasting to get a good understanding of this topic and address the importance and relevance of this research. As cognitive biases are an important aspect of judgmental forecasting that influences performance, this is the second topic of this literature review. As explained in Section 1.2, the segmentation categories are a well-known and essential topic in operations management. This is the third aspect of which the literature is reviewed. The last topic of the literature study is decisional guidance.

2.1 Judgmentally adjusting forecasts

Forecasting is one of the important processes in operations management. For companies, forecasting sales is about making accurate predictions of the actual sales of certain products over a certain period (Ritzman & King, 1993). Accurate sales forecasts can help companies with decision-making and reducing costs by means of the inventory holding costs (Ritzman & King, 1993). Furthermore, it can increase the service level and allow companies to consider changes in the economic environment. The sales forecasts also impact many decisions outside a company, in the entire supply chain (Fildes et al., 2006). This highlights the importance of an accurate forecast even more. Sales forecasting is the premise of strategic planning, decisions on ordering, and the distribution of orders. This entails that sales forecasting can improve the business's operational efficiency and customer satisfaction. Companies make their forecasts in multiple ways. It is done manually by a judgmental forecast of the demand planners, based on a statistical forecast or as a combination of the two previously named methods (Nahmias & Olsen, 2015). Many companies use the combination of statistical forecasts with judgmental forecasts, and it is the topic of many academic articles in the field of behavioral operations management and is referred to as integrated forecasting methods (Goodwin, 2002). The integrated forecast methods consist of two subcategories, integrated mechanical methods and voluntary integrated methods. Mechanical integration of forecast implies applying a statistical method combined with a judgmental forecast (Goodwin, 2002). Voluntary integrated forecasting methods combine judgmental forecasts of demand planners and statistical forecasts where planners can perform a judgmental adjustment on the statistical forecast (Arvan et al., 2019). For this research, the focus is on voluntary integrated forecasting methods. The process of judgmentally adjusting the forecast consists of two phases (Arvan et al., 2019).

First, the planner needs to decide whether an adjustment of the forecast generated by the system is necessary. When the planner has decided an adjustment is necessary, the second phase consists of deciding on the direction and the size of the adjustment.

The adjustments of statistical forecasts by planners are conducted in order to improve the forecast accuracy (Li & Jiang, 2017). Based on many researches done over the past decades, in many cases, judgmental adjustments by planners on statistical forecasts do not improve forecast accuracy. In the early days of judgmental forecasting, there was no clear acceptance of the importance of judgment in forecasting (Lawrence et al., 2006). Lawrence et al. (2006) indicated that in later stages of research about judgmental forecasting, a general consensus on the importance of judgments in forecasting had been addressed in academic papers by many authors. An experiment of Worthen (2003) clearly showed what could happen when companies left judgments out of all forecasting stages. Due to the inaccuracy of the statistical forecasts and the lack of existence of any judgment, the experiment of Nike resulted in a huge inventory write-off.

Many other articles show substantial evidence that including a judgmental adjustment focusing on special events not taken into account by the statistical forecast can improve the forecast accuracy (Fildes et al., 2009). The current issues and dynamics in the current business settings make it very difficult to only rely on statistical forecasting methods (Alvarado-Valencia et al., 2017). Many studies researched the perfect circumstances in which it is beneficial or necessary to adjust certain forecasts judgmentally. Arvan et al. (2019) state in their research that the conditions depend on the characteristics of a time series. A time series with a very high variability might degrade the efficiency of the statistical models. In such a case, a judgmental adjustment with information about certain events not taken into account in the statistical models can be relevant and useful to improve the forecast accuracy (Arvan et al., 2019). Regarding the judgmental adjustments a planner performs, Alvarado-Valencia et al. (2017) concluded that the expertise of the demand planner is a crucial aspect. The circumstances and usage of expert knowledge to perform a judgmental adjustment might affect the judgmental adjustments. In their review, Alvarado-Valencia et al. (2017) state the choice of method of elicitation of knowledge can affect the responses given by the planners. Furthermore, the number and selection of planners, the personal attributes of the planners and the way of demonstrating the information can affect the responses.

2.2 Cognitive biases in judgmental forecasting

The literature described in section 2.1 discussed some characteristics of planners that positively influence the performance of a planner in adjusting the forecast. However, there is also a significant amount of literature on the characteristics that could negatively affect the performance of a planner in adjusting the forecast. One of the main disadvantages of judgmental forecasting described in many studies is the fact that planners can be biased when making adjustments (Sanders & Ritzman, 2004).

Cognitive biases in judgmental forecasting can cause lower performance when adjusting statistical forecasts (Fildes et al., 2009). In many cases, literature on this topic focuses on the overall performance of the planners and the presence of biases. The research of Pennings (2016) focused on individual differences among planners instead of aggregate results in judgmental forecasting. This behavior is forecasting heterogeneity (Pennings, 2016). Focusing on individual differences and the individual presence of biases can help analyze possibilities for improvements in the judgmental forecasting process. Pennings (2016) investigated the differences between demand smoothers and demand chasers. Based on earlier articles of Kremer et al. (2011) and Lau et al. (2014), Pennings (2016) created a more advanced model to analyze the differences among demand smoothers and demand chasers. They found differences in the performance of the demand chasers versus the demand smoothers, as well as the presence of demand chasers or smoothers among different roles and departments. This means the environment and the role of a forecaster can impact the forecaster's behavior. Eroglu & Croxton (2010) explained another method to define individual differences. The research of Eroglu & Croxton (2010) focuses on the optimism bias, the anchoring bias and the overreacting bias. The optimism bias refers to the tendency to project mainly positive sales results in the future, which leads to mainly positive errors (Eroglu & Croxton, 2010). The anchoring bias implies the situation of a forecaster adjusting statistical forecasts in the right direction, but the adjusted value stays too close to the statistical forecast as an anchor value (Eroglu & Croxton, 2010). The overreaction bias implies when a planner adjusts in the right direction but overshoots the actual sales resulting in a too large adjustment that increases the error (Eroglu & Croxton, 2010). The formulas and logic behind these biases come from the research of Eroglu & Croxton (2010). The base of the explanation of the biases starts with the percentage errors of the forecasts. The following formulas of percentage error of an observation i have an important role in this thesis:

$$\text{Percentage error of statistical forecast: } p_{\mathrm{s}}^{\mathrm{i}} = 100\left(\frac{f_{\mathrm{s}}^{\mathrm{i}} - s^{\mathrm{i}}}{s^{\mathrm{i}}}\right) \tag{2.1}$$

$$\text{Percentage error of adjusted forecast: } p_{\mathrm{a}}^{\mathrm{i}} = 100\left(\frac{f_{\mathrm{a}}^{\mathrm{i}} - s^{\mathrm{i}}}{s^{\mathrm{i}}}\right) \tag{2.2}$$

with

s^{i} = actual demand size, $f_{\mathrm{s}}^{\mathrm{i}}$= statistical forecast, and $f_{\mathrm{a}}^{\mathrm{i}}$ = adjusted forecast

To be able to calculate the biases, three different formulas are used by Eroglu & Croxton (2010). In each formula, the variable n is present. The n indicates the total number of adjustments. The logic behind the optimism bias is based on the tendency of a planner to mainly focus on the positive results when forecasting, which results in more positive errors (Eroglu & Croxton, 2010). The formula uses the percentage errors of the adjusted forecasts. If there is no optimism bias, the amount of positive and negative forecast errors should be almost equal. Therefore, the formula of the optimism bias takes the average percentage error of all the n adjustments. To calculate a bias for a specific planner, only the n adjustments of that planner are included. If there is no optimism bias, the value should equal zero since there are equal positive and negative errors (Eroglu & Croxton, 2010).

$$\text{Optimism bias: } B_{\mathrm{o}}^{\mathrm{i}} = \frac{1}{n}\sum_{i=1}^{n} p_a^i \tag{2.3}$$

The anchoring bias implies the situation of a forecaster adjusting statistical forecasts in the right direction, but the adjusted value stays too close to the statistical forecast as an anchoring value (Eroglu & Croxton, 2010). In this situation, two variables are used, variable x and variable y. The variable x gets a value of 1 when the adjusted forecast improves the accuracy, and the adjusted forecast has the same sign as the statistical forecast (+ or -). The variable y gets a value of 1 for a lower forecast error due to the adjustment, and otherwise a zero. With these variables, the value for the anchoring bias is calculated. If there is no anchoring bias, the value should equal 0.5. A value of 0.5 means that from all the adjustments that improved the accuracy, in 50 % of them the adjusted forecast was between the statistical forecast and the actual demand. If the value is higher than 0.5, the planner tends to anchor more on the statistical forecast.

$$\text{Anchoring bias: } B_{\mathrm{a}}^{\mathrm{i}} = \frac{\sum_{i=1}^{n} x^i}{\sum_{i=1}^{n} y^i} \tag{2.4}$$

with

$$x = \begin{cases} 1 \text{ if } |p_{\mathrm{a}}^{\mathrm{i}}| < |p_{\mathrm{s}}^{\mathrm{i}}| \text{ and } p_{\mathrm{a}}^{\mathrm{i}} p_{\mathrm{s}}^{\mathrm{i}} > 0 \\ 0 \text{ otherwise} \end{cases} \tag{2.5}$$

$$y = \begin{cases} 1 \text{ if } |p_{\mathrm{a}}^{\mathrm{i}}| < |p_{\mathrm{s}}^{\mathrm{i}}| \\ 0 \text{ otherwise} \end{cases} \tag{2.6}$$

The overreaction bias implies when a planner adjusts in the right direction but overshoots the actual sales resulting in a too large adjustment increasing the forecast error (Eroglu & Croxton, 2010). The variable z is used for the overreaction bias. In this situation, the variable z gets a value of 1 when the judgemental adjustment results in a greater error, and the directions of the adjustment error and statistical error are opposite. In any other situation, the value becomes 0. The value of overreaction bias gets calculated by taking the average number of z of all the n adjustments. If there is no overreaction bias, the result of the formula should equal zero.

$$\text{Overreaction bias: } B_{\mathrm{r}}^{\mathrm{i}} = \frac{1}{n} \sum_{i=1}^{n} z^i \tag{2.7}$$

with

$$z = \begin{cases} 1 \text{ if } |p_{\mathrm{a}}^{\mathrm{i}}| > |p_{\mathrm{s}}^{\mathrm{i}}| \text{ and } p_{\mathrm{a}}^{\mathrm{i}} p_{\mathrm{s}}^{\mathrm{i}} < 0 \\ 0 \text{ otherwise} \end{cases} \tag{2.8}$$

The research of Eroglu & Croxton (2010) focuses on the effects of personality and motivational differences on forecasting biases. For the optimism bias, they concluded that the personality trait of openness to experience increases this bias, while agreeableness decreases the presence of the optimism bias. For the anchoring bias, several variables influence the existence of this bias, indicated by the research of Eroglu & Croxton (2010). People who are high in conscientiousness and agreeableness and low in extraversion score higher on the anchoring bias. Furthermore, challenge-seekers tend to score lower on the anchoring bias. The last bias is the overreaction bias. In this case, the personality traits external and internal work locus of control decrease the presence of the overreaction bias. Furthermore, the personality trait extraversion also decreases the presence of overreaction bias, while conscientiousness increases the bias. These findings lead to the conclusion that the presence of the three biases differs greatly depending on demand planners' personalities.

In the literature on personal differences in judgmental forecasting, more different methods are used to analyze how these differences affect forecasting accuracy. For example, the research of Moritz et al. (2014) investigated the effect of cognitive reflection on performance. In their study, they used the cognitive reflection test (CRT) of Frederick (2005) to indicate differences between the planners. Based on their study, they concluded that individuals that score higher on the CRT make more accurate forecasts.

2.3 Categorization in demand analysis

The importance of forecasting is evident in the entire supply chain. Forecasting accurate numbers of expected demand influences many other decisions in many aspects of the supply chain, such as inventory management (Fildes et al., 2009). The expected demand for a certain period influences a company's inventory management. The inventory management of a company or a supply chain decides the parameters based on the forecasted values. Therefore there often is a collaboration between those parts of a supply chain or company. For forecasting it is important to have background information about inventory management. One of the components of the background information is the importance of a certain demand item. This importance can be clarified by classification. In supply chain inventory management, it is common to classify the demand items (Scholz-Reiter et al., 2012). Classification of items can be done in multiple ways, but in many cases, the classification uses the ABC and XYZ analyses. The ABC analysis usually ranks the products based on the annual turnover, and the XYZ analysis ranks products based on the level of uncertainty of demand (Scholz-Reiter et al., 2012). The ABC analysis ranks the items based on the turnover of the last twelve months (Scholz-Reiter et al., 2012). A-items cover 0 to 80 percent of the annual turnover, B-items 80 to 95 percent and C-items cover the last 5 percent. These rankings indicate the importance of items for companies which is very useful in various company processes. The XYZ analysis ranks products based on how difficult it is to predict the demand for a certain item. This classification is often used in judgmental forecasting to understand better which items are most important for planners to adjust judgmentally. These items are ranked based on the coefficient of variation in the demand. The coefficient of variation is the fraction of the standard deviation and the mean. X-items have a coefficient of variation below 0.5, Y-items have a coefficient of variation between 0.5 and 1 and Z-items have a coefficient of variation above 1. With these parameters, the items can be classified based on consumer behavior.

2.4 Decisional guidance in forecasting

When people have the task of making decisions, decisional guidance can be provided by a decision support system (DSS) to influence or support the users in their decision-making process. A DSS emphasizes the effectiveness and efficiency of decision-making (Parikh et al., 2001). Effectiveness of decision-making involves identifying what is necessary for the decision-making process(Parikh et al., 2001). Furthermore, it implies making sure the criteria chosen in the decision-making process should be relevant. The efficiency of decision-making implies the minimization of costs, effort and time used in the decision-making process (Montazemi et al., 1996).

Decisional guidance is how a DSS enlightens the decision-makers with structuring and executing the decision-making processes (Silver, 1991). A DSS can provide decisional guidance on purpose, deliberate guidance, or when unintentionally, inadvertent guidance (Silver, 1991). Since deliberate guidance is one that intends to influence the decision-maker, this is the main topic studied in earlier literature. Deliberate guidance can be divided into suggestive and informative guidance (Montazemi et al., 1996; Fildes et al., 2006). Suggestive guidance implies proposing action to the decision maker, while informative guidance gives unbiased, relevant information to the decision maker without any suggestions on actions to take. The effectiveness of the decisional guidance depends on the type of guidance in combination with the task and the goal of the guidance. Decisional guidance helps to structure the decision-making task and to execute the task (Silver, 1991). Guidance for structuring the process focuses on the choice users have to make in which operates to use and in which order they are used. Guidance for the execution of the decision-making process focuses on how the users perform regarding the evaluative and predictive judgments in the process. The research of Montazemi et al. (1996) found that suggestive guidance outperformed informative guidance when completing a less complex task. With a more complex task, participants receiving informative guidance outperformed those receiving suggestive guidance according to the study of Montazemi et al. (1996).

Regarding the performance or effectiveness of decisional guidance, there are quite some researches and case studies that studied the precise effects different types of guidance have on the decision-making process. Evaluating whether decisional guidance is effective has been measured on different variables in the literature. Parikh et al. (2001) used four different measurement variables to study the effects of decisional guidance: Decision quality, satisfaction, learning and efficiency. Parikh et al. (2001) found suggestive guidance to be more effective in improving decision quality and user satisfaction. Informative guidance came out to be more effective regarding user learning. Both informative and suggestive guidance reduced the decision time and therefore improved the decision-making efficiency.

According to Silver (1991), decisional guidance is most useful in a system that is not very restrictive since there will not be much opportunity for guidance in a very restrictive system. This also means, in general, for each judgmental opportunity in a system, during the design phase must be decided whether there will be guidance for the process, do nothing, or restrict the process (Silver, 1991). The design of a DSS depends mainly on three elements (Parikh et al., 2001). The first element is the task, in which the most important characteristics are the type, structure, frequency and complexity of the task. Secondly, the user and especially its characteristics, capabilities and needs is an important element. The last element is the context of the organization, implying the level in the organization where the decision will be taken, the situation and the purpose of the decision.

The classification of decisional guidance has been done by Silver (1991) on targets, forms, scopes and modes. Besides the different forms of decisional guidance, deliberate and inadvertent guidance, there are also three modes (Silver, 1991). The first mode is predefined guidance, in which the system's recommendations and displays with information are designed beforehand completely by the designer. Dynamic guidance is the mode in which the recommendations and displays are generated by the system in a dynamic way. This means the design of the guidance has been made beforehand and completed with inputs from the systems like numbers and other relevant data. The last mode is participative guidance, in which users need to interact with the system in a participative way to determine the content and format of the guidance that will be provided.

There has been quite some research on decisional guidance in general, but research on decisional guidance in judgmental forecasting is limited. Fildes et al. (2006) conducted research on the design of a Forecasting Support System (FSS) and how it should look in the ideal situation. An FSS is a particular type of DSS. A support system's task has been defined as a judgment of a manager plus the system's model in order to provide an effective solution (Keen, 1978). According to the article of Fildes et al. (2006), the key features of an FSS in integrating the model and the judgment are the database, a set of quantitative forecasting techniques and applications that give the possibility to managerial judgments. To be able to conclude the characteristics of an ideal FSS, Fildes et al. (2006) first analyzed the supply chain forecasting task. To make an accurate forecast, relevant data must be accessible in the FSS. In the paper of Fildes et al. (2006), they described six different types of data. The time series data is the first type of data. Time series data can be at multiple different levels of aggregation, such as product group or family, region and country. There are three typical components in time series: regular patterns, irregular components from predictable events and noise (Fildes et al., 2006). Earlier made forecasts in other periods and forecasts made in the same period are also defined as types of data that should be made available in an FSS. Related to earlier forecasts, information on errors made in these forecasts should also be accessible such that it can be provided as feedback to the planners. The last two types of data described by Fildes et al. (2006) are information on customers' activities and information on other relevant variables. Examples of other relevant variables are weather forecasts and the special activities of competitors. In the ideal situation of an FSS, the statistical model used to generate the forecast should clarify all regular patterns (Fildes et al., 2006). Furthermore, in this ideal situation, the irregular but foreseeable events should be clarified by the planner with a judgmental forecast. However, as stated in earlier sections, this ideal situation is not the reality in many situations since planners also adjust based on patterns that are already included in the statistical method or they do not value the relevant information correctly (Fildes et al., 2006, 2009; Arvan et al., 2019; Lawrence et al., 2006).

To be able to tackle the inefficiency and ineffectiveness of judgmental adjustments, an ideal FSS should be able to improve the planner's judgment on when an adjustment is necessary. Furthermore, the ideal FSS should enable the planner to make accurate judgmental adjustments when this is necessary. Decisional guidance in an FSS can have the same types of forms as described before, informative guidance and suggestive guidance. Informative guidance that can be valuable within forecasting is feedback (Fildes et al., 2006). Feedback can be given in multiple ways, such as simply the latest outcome defined as outcome feedback (Benson & Önkal, 1992). Other types of feedback can be performance feedback, cognitive process feedback, or task properties feedback. Performance feedback implies providing information on the forecast accuracy and cognitive process feedback implies providing specific information to the planner about his or her strategy. With task properties feedback, the user gets statistical information about the task (Benson & Önkal, 1992). Informative guidance can improve the decision-making process in general (Montazemi et al., 1996). The other form of decisional guidance is suggestive guidance. Suggestive guidance implies directly suggesting actions in order to give advice (Fildes et al., 2006). The general effectiveness of suggestive guidance has also been proved in multiple studies (Montazemi et al., 1996; Parikh et al., 2001). However, Fildes et al. (2006) described that only suggestive guidance without any explanation of the advice could lead to miscalibration. The article of Fildes et al. (2006) suggests guidance rather than restrictiveness. Restrictiveness in FSS implies limiting the user in the amount of data shown or in the available views (Fildes et al., 2006). It can also restrict certain actions in the process. According to the article of Fildes et al. (2006), the use of absolute restrictiveness can be very dangerous since it might be frustrating to the users and it can be difficult for the designer to determine which processes can be the most relevant to use. Regarding the form of decisional guidance that would be best, Singh (1998) concluded a combination of suggestive and informative guidance led to better decision-making.

Chapter 3

Data analysis

This research aims to get insights into the current performance of the different planner types and to improve the guidance planners receive in the decision-making process of judgmental adjustment. To answer the first research question, data is collected and analyzed to have a good overview of the current situation and performance of the planners. This analysis shows where the performance of the adjustments is lacking, and thus, decisional guidance can be used in order to steer the planner in such a way that the accuracy of the judgmental forecasts will increase. We analyze the performance of the planners among the different segmentation categories to see if it is clear on which categories more guidance is needed. Before analyzing the data, the data needs to be cleaned.

3.1 Data and filtering

The dataset originates from a customer of EyeOn, company A. The dataset needs some filtering steps before it can be used to perform the analysis. Earlier research within EyeOn and in the literature showed that planners, in general, have the ability to make a good judgment about whether an adjustment is needed but often make mistakes regarding the right size and/or direction (Cuppens, 2020; Sanders & Ritzman, 2004; Fildes et al., 2009). For this research, the adjustments done at Company A over the demand for items in previous months will be used. To draw conclusions about this, the dataset needs to meet certain requirements. At EyeOn, the adjustment data for company A has been logged for a longer period, but since February 2022 EyeOn changed the method of logging it in order to be able to see on which level the adjustment has been made. For the analysis regarding the different segmentation categories, it is important that the adjustments are made on product level. The segmentation with the ABC/XYZ ranking is on product level, so the adjustments also need to be made on product level. For the analysis of the planner differences, it is not necessary that the adjustments are made on product level. This means that the different analyses also will have different data.

Another point that is filtered on is the adjustment size. The adjustment size is the difference between the final forecast and the forecast before. If the adjustment size equals 0, these adjustments have been removed from the dataset since this means there has been no actual adjustment made. Another step in the filtering process is removing all adjustments with a statistical forecast of 0. Statistical forecasts of 0 often have been removed since usually there are specific reasons or agreements between EyeOn and the customers why these forecast items do not have a forecast. An explanation can be that there is too little information for the system to generate a good forecast and therefore, a manual forecast made by a planner is required. This is one of the situations where a planner adds value, but in this case, this is not an adjustment, and therefore they are not interesting for the scope of this research. Another occasion when adjustments are removed is when the final forecast or the actual demand is 0. When the final forecast is zero after an adjustment, the planner probably has specific info directly from the customer why the demand will be zero. This is also a situation where the added value of a planner is visible. However, since the goal of this research is to focus on where decisional guidance can add value in the adjustment process of a planner focusing on the size of the adjustment, this situation does not fit with the scope of this research. As said before, adjustments where the forecast item has an actual demand of 0 also have been removed from the dataset used for this research. This decision was made since forecast items with an actual demand of 0 often have very different characteristics compared to the adjustments that are the main focus of this research. Furthermore, the categories EOL (end-of-life) and NPI (new product introduction) are left out of this analysis since forecasts for items in that category are made manually so there is no statistical forecast to adjust on. The data is filtered in Dataiku since this is the main program used within EyeOn for these purposes. After the filtering process, the adjustment overview data will be joined with the data of the segmentation categories and data of the actual demand. This is also done in Dataiku. Furthermore, all data from company A has been anonymized in this research, and the hierarchy levels are transformed into numbers. The statistical tests are performed in Dataiku and SPSS. For this data analysis, the t-tests are used to indicate statistical differences. There are no clear hypotheses in this part, and the tests are only used to identify interesting differences regarding the performance of the judgmental adjustments. For the performance of the adjustments among the different segmentation categories, unpaired two-sample t-test are used since the comparison is made between different observations of adjustments. For the biases one sample student t-tests are used to test if the biases are significantly different from 0 (optimism bias and overreaction bias) and 0.5 (anchoring bias).

Before the data can be used for analysis, the dataset needs to be checked for outliers. Outliers in datasets differ significantly from other data points. These outliers can have significant effects on the analyzes and insights retrieved from the analyses. Outlier correction can be done based on multiple methods, such as the Inter Quartile Range or the usage of z-scores. For this thesis, the Inter Quartile Range is the method for the outlier correction with a lower and upper bound of 1.5. For this specific case, the errors of the statistical and adjusted forecast are used as variables. If one of the errors is an outlier, the adjustment is removed from the dataset. Removing the outliers from the dataset affects the results of the analysis. If the outliers are not removed, some really large errors can have a huge impact on the results while such an outlier might not be a good representation of the entire dataset.

Below in Table 3.1 the details about the data preparation process are visible. Table 3.2 shows the distribution among the different business units.

3.2 Measuring accuracy

There are many different accuracy measurement methods known in forecasting literature. In this section, a few of them are explained. Using different measurements can lead to different conclusions (Davydenko & Fildes, 2013). A well-known error measurement in

Table 3.1: Statistics Dataset

Adjustments	# Observations
Before data cleaning	12,353
Missing values	1,300
EOL or NPI	35
Statistical forecast of 0	255
Size of 0	2747
Forecast after of zero	237
Actual demand quantity of 0	3,918
After data cleaning	3,861
After outlier correction	3,062

Table 3.2: Statistics BU's

	# Observations
All adjustments	3,062
Adjustments from BU D	2,016
Adjustments from BU E	767
Adjustments from BU F	279

forecasting is the Mean Absolute Error (MAE). The MAE uses the absolute difference between the forecast and the actual demand. The MAE is the average of all the errors. Another method is the mean squared error (MSE). The MSE squares each error and then takes the mean of all squared errors. This different approach results in the MSE penalizing large absolute errors more than the MAE. Although these methods are known methods, over the last decades in literature, most researchers did not use these methods anymore (Hyndman & Koehler, 2006; Davydenko & Fildes, 2013). Another common method is the mean absolute percentage error (MAPE). The MAPE is commonly used to compare forecasts (Bowerman et al., 2005). It takes the absolute percentage error of each observation, and the mean of all of them gives the MAPE. A disadvantage of using the MAPE is the fact that it is sensitive to outliers, zero values and values that are close to zero. A low actual value and a high forecast results in a very high absolute percentage error. However, since the fact that zero values are removed from the dataset and outlier detection is used in combination with the fact that the MAPE is the metric that is used within EyeOn, the MAPE is the main metric for this thesis. Furthermore, the differences in performance among different planner types will be analyzed, also the performance over the different segmentation categories. The Forecast Value Add (FVA) is the most important aspect since it shows the actual value that has been added by that actual adjustment. The MAPE might give an incomplete view of the situation since it is more important to see what a certain action of a planner added to the performance.

As explained in Section 1.5, in the process at EyeOn, it is possible to adjust an item multiple times in each planning period. Therefore, there is a statistical forecast, a forecast before (the adjustment) and a forecast after (the adjustment). To get a good overview of the performance of an adjustment, it is important to analyze the difference between the forecast after and the forecast before the adjustment which is done with the $FVA^{\mathrm{i}}_{\mathrm{b,a}}$.

To get a good overview of the entire situation, the performance of the statistical forecast is also included. Below the formulas are visible. Since the process at EyeOn is a bit different, the formulas differ a bit from the earlier shown formulas in Section 2.2.

$$\text{Percentage error of statistical forecast: } p_{\mathrm{s}}^{\mathrm{i}} = 100\left(\frac{f_{\mathrm{s}}^{\mathrm{i}} - s^{\mathrm{i}}}{s^{\mathrm{i}}}\right) \tag{3.1}$$

$$\text{Percentage error of forecast before: } p_{\mathrm{b}}^{\mathrm{i}} = 100\left(\frac{f_{\mathrm{b}}^{\mathrm{i}} - s^{\mathrm{i}}}{s^{\mathrm{i}}}\right) \tag{3.2}$$

$$\text{Percentage error of forecast after: } p_{\mathrm{a}}^{\mathrm{i}} = 100\left(\frac{f_{\mathrm{a}}^{\mathrm{i}} - s^{\mathrm{i}}}{s^{\mathrm{i}}}\right) \tag{3.3}$$

with

s^{i} = actual demand size, $f_{\mathrm{s}}^{\mathrm{i}}$= statistical forecast, $f_{\mathrm{b}}^{\mathrm{i}}$ = forecast before, and $f_{\mathrm{a}}^{\mathrm{i}}$ = forecast after

$$\text{MAPE of the statistical forecast: } MAPE_{\mathrm{s}}^{\mathrm{i}} = \frac{1}{n} * \sum_{i=1}^{n} |p_{\mathrm{s}}^{\mathrm{i}}| \tag{3.4}$$

$$\text{MAPE of the forecast before: } MAPE_{\mathrm{b}}^{\mathrm{i}} = \frac{1}{n} * \sum_{i=1}^{n} |p_{\mathrm{b}}^{\mathrm{i}}| \tag{3.5}$$

$$\text{MAPE of the forecast after: } MAPE_{\mathrm{a}}^{\mathrm{i}} = \frac{1}{n} * \sum_{i=1}^{n} |p_{\mathrm{a}}^{\mathrm{i}}| \tag{3.6}$$

$$\begin{gathered}\text{FVA of statistical forecast compared to forecast after:}\\ FVA_{\mathrm{s,a}}^{\mathrm{i}} = MAPE_{\mathrm{s}}^{\mathrm{i}} - MAPE_{\mathrm{a}}^{\mathrm{i}}\end{gathered} \tag{3.7}$$

$$\begin{gathered}\text{FVA of forecast before compared to forecast after:}\\ FVA_{\mathrm{b,a}}^{\mathrm{i}} = MAPE_{\mathrm{b}}^{\mathrm{i}} - MAPE_{\mathrm{a}}^{\mathrm{i}}\end{gathered} \tag{3.8}$$

Chapter 4

Results data analysis

In this section, the results of the data analysis will be shown and explained. The data analysis is performed on a dataset of adjustments made by planners from a customer of EyeOn. The goal of this section is to get insights into the behavior of planners. This will be analyzed based on the current performance, and the different decisions planners take during the adjustment process.

4.1 Metrics

The main metrics that are being compared with each other in this section are the MAPE, the FVA, the optimism bias, the anchoring bias and the overreaction bias. The outcomes of these measures are used to show whether there are differences in the metrics split by segmentation categories, hierarchy levels and the different planners present in the dataset. The most important focus is on the FVA since the FVA gives a complete view of the situation. The MAPE is also shown so that it is clear where the FVA comes from. To test the effect of the adjusted forecasts, student t-tests are performed. The student t-test is performed to see if planners have significant biases. This means that for the individual planners, the value of the corresponding bias is significantly different from 0 (optimism bias and overreaction bias) or 0.5 (anchoring bias). Furthermore, descriptive statistics about other interesting findings are presented in this section. Below in Table 4.1 some explanation of the used metrics and abbreviations is given.

4.2 General Results

Table 4.2 shows general results of the performance of the statistical forecast ($\mathrm{MAPE_a}$), the forecast before ($\mathrm{MAPE_b}$), the forecast after ($\mathrm{MAPE_a}$) and the corresponding FVA's.

Based on the general results, it is clearly visible that the current process of judgmental forecasting is far from optimal. The forecast value add going from the statistical forecast to the adjusted forecast, $\mathrm{FVA_{s,a}}$ is negative, i.e., the forecast error increases. Since in the way of working adjusting forecasts in this case, there can be multiple adjustments on the same item in each planning version, comparing these forecasts gives other results. The $\mathrm{FVA_{b,a}}$ compares to adjusted forecast with the forecast before that adjustment. This will be equal to the $\mathrm{FVA_{s,a}}$ if it is the first adjustment for that item in the current planning version. If it is not the first adjustment, the value of the $\mathrm{FVA_{b,a}}$ will be different from the $\mathrm{FVA_{s,a}}$. The results show that compared to the FFB the adjustment the adjustments

Table 4.1: Explanation metrics & abbreviations

Metric	Explanation
f_{s}	Statistical forecast
f_{b}	Forecast before adjustment
f_{a}	Forecast after adjustment
$\mathrm{MAPE_s}$	Mean Absolute Percentage Error of f_{s}
$\mathrm{MAPE_b}$	Mean Absolute Percentage Error of f_{a}
$\mathrm{MAPE_a}$	Mean Absolute Percentage Error of f_{b}
$\mathrm{FVA_{s,a}}$	Forecast value added with f_{a} compared to f_{s}
$\mathrm{FVA_{b,a}}$	Forecast value added with the f_{a} compared to the f_{b}
$\mathrm{B_o}$	Optimism bias
$\mathrm{B_a}$	Anchoring bias
$\mathrm{B_r}$	Overreaction bias

Table 4.2: General results

Metric	Overall result
$\mathrm{MAPE_s}$	46.20%
$\mathrm{MAPE_b}$	54.88%
$\mathrm{MAPE_a}$	52.92%
$\mathrm{FVA_{s,a}}$	-6.72%
$\mathrm{FVA_{b,a}}$	1.96%

do add value with an $\mathrm{FVA_{b,a}}$ of 1.96%. This means, on average, that the first adjustments in a planning version are the worst and a major reason why the $\mathrm{FVA_{s,a}}$ of the adjustments is negative in this specific case. Table 4.3 gives an overview of the results of the first adjustments, adjusting the statistical forecast, and an overview of the results when excluding the first adjustments. All first adjustments in a planning version lead to an average $\mathrm{FVA_{s,a}}$ of -7.40% compared to an average of -6.72% of all adjustments. To get a good overview of the total dynamics in the adjustment process, Table 4.3 also gives the results of the direction choice of the planners. Since planners can make multiple adjustments to each version, it is interesting to see where and how often they make a choice for the right direction.

A closer look at Table 4.3 shows us that the majority (52.91%) of the first adjustments improve the statistical forecasts with an average $\mathrm{FVA_{s,a}}$ 30.55 % from a statistical MAPE of 56.36% to an average MAPE of 25.75 %. If there was no improvement of the statistical forecast, the $\mathrm{FVA_{s,a}}$ was on average -37.65%, from a statistical MAPE of 37.82% towards 75.47%. This gives the insight that planners also make mistakes in choosing whether an adjustment is necessary. However, the numbers of right direction versus the previous forecast show that, on average, planners are quite good at deciding whether an adjustment is necessary and especially in determining the direction of the adjustment. This insight also highlights the importance of investigating the opportunity of providing more guidance to planners in the adjustment process, focusing on the adjustment size.

Table 4.3: General results first adjustments

	$\text{FVA}_{\text{s,a}}$	$\text{FVA}_{\text{b,a}}$	Improved vs f_{b}	Improved vs f_{s}	Right direction vs f_{b}	Right direction vs f_{s}
First adj	-7.40%	-7.40 %	52.91 %	52.91 %	68.64 %	68.64 %
Excl first adj	-6.30%	7.60 %	58.14 %	40.82 %	73.05 %	55.73 %
Total	-6.72%	1.96 %	56.17 %	45.36 %	71.39 %	60.58 %

4.2.1 Direction

Table 4.4 gives more insights into the behavior of the planners regarding deciding the direction. With the knowledge from Table 4.3, that in most cases, planners adjust in the right direction, it is interesting to see the impact of choosing the right or wrong direction on the FVA's. Choosing the wrong direction leads to a larger negative FVA than the positive FVA of choosing the right direction. Choosing the wrong direction compared to the forecast before leads to an average $\text{FVA}_{\text{b,a}}$ of -42.45% while choosing the right direction only leads to an $\text{FVA}_{\text{b,a}}$ of 19.76%. When looking at the comparison with the SF, the same pattern is visible. Choosing the wrong direction leads to an $\text{FVA}_{\text{s,a}}$ of -30.97% while selecting the correct direction only leads to an $\text{FVA}_{\text{s,a}}$ of 9.06%. This highlights that there is a lot to gain regarding the decision on the adjustment size. While planners adjust in most cases in the right direction, they do not obtain much value out of choosing the right direction looking at the FVA's. For the experiment, this gives the insight that it can be valuable to focus on providing decisional guidance to planners when deciding on the size of the adjustment.

Table 4.4: Wrong or right direction

Situation	Frequency in %	$\text{FVA}_{\text{s,a}}$	$\text{FVA}_{\text{b,a}}$
Right direction vs f_{b}	71.39 %	4.58%	19.76%
Wrong direction vs f_{b}	28.61 %	-34.91%	-42.45%
Right direction vs f_{s}	60.58 %	9.06%	12.92%
Wrong direction vs f_{s}	39.42 %	-30.97%	-14.88%

4.2.2 Hierarchy levels

As explained in Section 1.5, each business unit has its own hierarchy logic. Therefore, the hierarchy levels are numbered in which hierarchy level 6 is always the lowest level on which the segmentation has been done. Looking at the differences among performance in different hierarchy levels can give insights into where planners perform better when adjusting forecasts. Details about the different hierarchy levels and the number of observations can be found in Table 4.5.

For this analysis, the metrics $\text{FVA}_{\text{s,a}}$, $\text{FVA}_{\text{b,a}}$ and the MAPE of the adjusted forecast will be analyzed in hierarchy level 6 compared to the average value of the metrics in the higher hierarchies. This will be done with paired student t-tests; the results are visible below.

The results of the analysis on the results of the different hierarchy levels show that planners have a significantly lower error on items of the lowest hierarchy level. This can

Table 4.5: Statistics hierarchy levels

Hierarchy level	# adjustments
1	444
2	126
3	221
4	249
5	375
6	279
All	1665

Table 4.6: Performance among hierarchy levels

Metric	HL 1-5	HL 6	t-statistic	p-value
$FVA_{s,a}$	-7.80%	-5.80 %	1.00	.31
$FVA_{b,a}$	3.13%	0.98%	-1.06	.29
$MAPE_a$	57.79%	48.83%	-4.92	<.001***

$* : p<.05, ** : p<.01, *** : p<.001$

be seen as a logical result since the lowest hierarchy level adjustments only contain an individual item which makes it easier to give a more precise adjustment. To be able to draw conclusions on the actual value that a planner adds with the adjustment, the $FVA_{s,a}$ and the $FVA_{b,a}$ are more interesting to analyze. From Table 4.6 can be concluded that the $FVA_{s,a}$ is higher for the lowest hierarchy level, but this is not the case for $FVA_{b,a}$. However, both measurements are not significantly different according to the student t-test. These results give the insight that it in practice it is relevant to adjustments made on all hierarchy levels could use guidance for planners to improve the adjustments.

4.3 Segmentation categories

In the dataset, roughly half of the adjustments are made on hierarchy level 6, corresponding to the level on which the segmentation categorization has been made. Below in Table 4.7, there is an overview of the number of adjustments made per different segmentation categories. It is clear and logical that there are not that many adjustments in the categories CX, CY, CZ and BX, which is in line with the guidelines belonging to the segmentation categories mentioned before in Section 1.5. However, since these numbers do not provide information about the percentages of the items in a category that have been adjusted, no clear conclusions can be drawn about whether planners made too many adjustments in a certain category.

Table 4.8 shows the general results on the performance of judgmentally adjusting forecasts among the items in ABC categories. From this result, it is visible that the average error is the lowest in items that belong to category A. However, this does not mean that planners add the most value over there, which is clearly visible looking at the $FVA_{s,a}$ and $FVA_{b,a}$ over the different categories. Planners perform the worst in category A looking at the $FVA_{s,a}$ and $FVA_{b,a}$. However Table 4.9 shows no significant differences visible in the $FVA_{b,a}$ scores of items among ABC categories. Planners make better forecasts in category A but add more value to the forecasts in categories B and C. This is also since

Table 4.7: Adjustments for each segmentation category

	X	Y	Z	Total
A	250	796	81	1127
B	33	289	130	452
C	1	26	55	82
Total	284	1111	266	1661

the statistical forecasts perform better for the products in a higher category, meaning category A outperforms category B and category B outperforms category C. Concluding, there are no significant differences found regarding the FVA among the different categories A, B and C. A reason for this could be that category A items are seen as more important and, therefore, have more advanced or multiple forecast techniques used for these items when generating the statistical forecasts.

Table 4.8: Performance among ABC categories

Metric	A	B	C
$MAPE_s$	40.80 %	46.16 %	57.25 %
$MAPE_b$	47.60%	53.72%	59.06 %
$MAPE_a$	47.25%	52.12%	53.33 %
$FVA_{s,a}$	-6.45%	-5.95 %	3.92 %
$FVA_{b,a}$	0.36%	1.60%	5.73 %

Table 4.9: Results $FVA_{b,a}$ differences ABC

Metric	A vs B	B vs C	A vs C	A vs BC
t-statistic	-0.42	-0.61	-0.91	-0.67
p-value	.68	.55	.36	.50

Besides the ABC categories, also the XYZ classification is a metric that EyeOn uses to classify forecast items. Since this classification identifies the uncertainty of demand, which can be seen as the difficulty in predicting the demand, more clear differences among the added value of planners are expected in this analysis. Below in Table 4.10, the general results of the MAPE and FVA scores among the different XYZ categories can be found.

The results from Table 4.10 show that planners add more value with their adjustments of items in category Z compared to items in categories X and Y. The error of adjusted forecasts increases with the level of uncertainty of demand (XYZ), which is in line with expectation because these items are more difficult to forecast. The average error of the adjusted forecasts is higher in category Z compared to X and Y. However, in category Z, the most value is added with the adjustments. This is also visible in Table 4.11, which shows the significant difference between the $FVA_{b,a}$. The performance of the statistical forecasts causes this difference. X and Y items have a more stable demand pattern, which makes it easier to forecast those items. That is visible in a lower error of the statistical

H]

Table 4.10: Performance among XYZ categories

Metric	X	Y	Z
$MAPE_s$	29.05 %	43.85 %	54.77 %
$MAPE_b$	38.65 %	49.72 %	62.24 %
$MAPE_a$	41.78%	50.86%	48.12 %
$FVA_{s,a}$	-12.73%	-7.01 %	6.65 %
$FVA_{b,a}$	-3.13%	-1.14%	14.12 %

forecasts for X and Y items compared to Z items. Since the $MAPE_s$ is already quite low in these categories, it is hard for planners to improve these forecasts, and that is visible in the results. Focusing on the experiment, this means it is more interesting for X and Y items to guide planners not to adjust while guiding Z items on the size of the adjustment.

Table 4.11: Results $FVA_{b,a}$ differences XYZ

Metric	X vs Y	Y vs Z	X vs Z	XY vs Z
t-statistic	-0.56	-4.11	-4.15	-4.41
p-value	.58	<.001***	<.001***	<.001***

*:p<.05, **:p<.01, ***:p<.001

To be able to gain more insights into the different behavior of planners Tables 4.12,4.13, 4.14 and 4.15 provide more specific insights on where planners add value. The tables show that planners are quite good at choosing the direction of an adjustment in all categories, but planners are better at choosing the right direction of an adjustment for items from category Z compared to items from categories X and Y. While planners often select the right direction (68,39% of the adjustments on the lowest hierarchy level), not all of those adjustments in the right direction lead to improved forecasts. This is because planners adjust in the right direction but adjust the forecast too heavily and overshoot the actual demand. The result of this action is visible in the Tables 4.12,4.13, 4.14 and 4.15 and will be further analyzed in 4.6. What can be concluded from the tables below is the fact that planners can definitely improve in the process of choosing the right size for the adjustment. Since it was visible before in Table 4.10 that the statistical forecasts of items of category X and Y have a lower error than items from category Z. However, for all categories, the majority of the adjustments are in the right direction. This makes it interesting to test if it works better for items in categories X and Y to focus on guidance on the decision of making an adjustment or only on the size of the adjustment.

Table 4.12: Right direction vs f_b

	X	Y	Z	Total
A	64.40%	67.34%	82.72%	67.79%
B	69.70%	67.13%	76.15%	69.91%
C	100%	76.92%	81.82%	79.27%
Total	64.79%	67.51%	79.32%	68.39%

Table 4.13: Improve vs f_s

	X	Y	Z	Total
A	50.40%	54.40%	70.37%	54.66%
B	54.55%	51.90%	62.31%	55.09%
C	100%	50.00%	60.00%	56.10%
Total	50.70%	53.56%	64.29%	54.85%

Table 4.14: Right direction vs f_s

	X	Y	Z	Total
A	60%	62.31%	75.31%	62.73%
B	60.61%	61.25%	70.00%	63.72%
C	0.00%	84.62%	76.36%	78.05%
Total	59.86%	62.56%	72.93%	63.76%

Table 4.15: Improve vs f_b

	X	Y	Z	Total
A	41.60%	49.37%	61.73%	48.54%
B	54.55%	48.79%	59.23%	51.55%
C	0.00%	53.85%	56.36%	54.88%
Total	41.90%	53.56%	49.32%	49.67%

4.4 Optimism bias

The optimism, anchoring and overreaction bias have been calculated with the method used by Eroglu & Croxton (2010), explained in Section 2.2. The optimism bias can be calculated with the average percentage error of the adjusted forecast. When there is no optimism bias present, this number should be around zero. If the number is significantly higher than zero this would mean there is optimism bias. To identify the behavior of the planners, for all planners a students t-test will be conducted to test whether a planner can be identified with the optimism bias. The biases are tested on 10 planners with sufficient amount of adjustments (minimum of 40). This minimum of 40 is chosen since there was a clear gap between the number of observations below 40.

Table 4.16: Results Optimism bias

Planner	# Adjustments	B^o	t-statistic	p-value
A1	1430	0.17	10.03	<.001***
A3	299	0.09	2.22	.03*
B1	244	0.40	8.04	<.001***
B3	240	0.07	1.32	.19
A2	220	0.12	2.48	.01*
B2	163	0.29	4.67	<.001***
A4	67	0.08	0.87	.39
C1	65	0.14	1.94	.06
B7	64	-0.44	-7.19	<.001***
C2	40	0.30	2.26	.03*
Total	3062	0.16	12.82	<.001***

*:p<.05, **:p<.01, ***:p<.001

A score of zero on the optimism bias would mean a planner does not have an optimism bias and has almost an equal amount of negative errors as positive errors. From the results in Table 4.16 it is visible that six planners have a significant optimism bias. The average score of the optimism bias is 0.16, and from the table, it is visible that especially planners A3, B2 and C2 score way higher than this average score. Furthermore, it is interesting to see that planner B7 has a negative score instead of an expected positive score. This means this planner has way more negative errors.

4.5 Anchoring bias

The formula for calculating the anchoring bias has also been retrieved from Eroglu & Croxton (2010). The anchoring bias is calculated based on the fraction of all adjustments that lead to an improvement but where the adjusted value stayed too close to the statistical or anchor value. On average, there should be around the same amount of adjustments that improve the accuracy and have a positive error (final forecast higher than actual demand) as improved adjustments with a negative error (final forecast lower than actual demand). This would mean the score of anchoring bias would be around 0.5. As explained earlier in this section, there are two values planners can see as input for their adjustment. This can be the statistical forecast or the forecast before. In the analysis, both values are used to calculate an anchoring bias. This gives two tables, Table 4.17 about the anchoring on the statistical forecast and Table 4.18 about the anchoring on the forecast before.

Table 4.17: Results Anchoring bias vs f_{s}

Planner	# Adjustments	B^a	t-statistic	p-value
A1	1430	0.58	4.65	<.001***
A3	299	0.65	4.03	<.001***
B1	244	0.47	-0.46	.65
B3	240	0.70	3.74	<.001***
A2	220	0.61	2.31	.02*
B2	163	0.59	1.49	.14
A4	67	0.63	1.66	.11
C1	65	0.76	2.75	.01*
B7	64	0.00	0.00	.00
C2	40	0.60	0.76	.46
Total	3062	0.61	8.04	<.001***

*:p<.05, **:p<.01, ***:p<.001

Table 4.18: Results Anchoring bias vs f_{b}

Planner	# Adjustments	B^a	t-statistic	p-value
A1	1430	0.60	5.47	<.001***
A3	299	0.70	5.57	<.001***
B1	244	0.47	-0.83	.41
B3	240	0.64	3.41	<.001***
A2	220	0.62	2.40	.02*
B2	163	0.61	2.53	.01*
A4	67	0.64	1.91	.06
C1	65	0.82	4.88	<.001***
B7	64	0.83	4.59	<.001***
C2	40	0.63	1.15	.26
Total	3062	0.62	9.81	<.001***

*:p<.05, **:p<.01, ***:p<.001

Tables 4.17 and 4.18 show that there are five planners (A1, A3, B3, A2 and C1) that show a significant anchoring bias on both the statistical forecast and the forecast before. Planners B2 and B7 only have a significant anchoring bias on the forecast before. Planner B7 does not have any score in Table 4.17, meaning there are no adjustments in which any anchoring on the statistical forecast took place. This is interesting since planner B7 does show a significant high anchoring bias on the forecast before. Planner B1 has the score closest to having no anchoring bias at all in both situations.

4.6 Overreaction bias

The formula for calculating the overreaction bias has also been retrieved from Eroglu & Croxton (2010). In this formula, the sum of all adjustments in which a planner overreacts gets divided by the total number of adjustments. An overreaction is an adjustment in the right direction, but the planner overshoots the actual demand such that the forecast error increases. For the overreaction bias, the main goal of the analysis is to identify whether the overreaction bias is present in the planners and to what extent. Also, for the overreaction bias, there are two values used for calculating this bias, the statistical forecast and the forecast before.

Tables 4.19 and 4.20 show that almost all planners have a significant overreaction bias. The results of the planners are quite similar to the statistical forecast versus the forecast before. Planners B2 and C2 show quite a big difference between overreaction versus the statistical forecast compared to the forecast before.

Table 4.19: Results Overreaction bias vs f_s

Planner	# Adj	B^r	t-statistic	p-value
A1	1430	0.14	15.24	<.001***
A3	299	0.13	6.78	<.001***
B1	244	0.19	7.51	<.001***
B3	240	0.18	7.12	<.001***
A2	220	0.18	6.59	<.001***
B2	163	0.29	8.22	<.001***
A4	67	0.09	2.55	.01*
C1	65	0.20	4.00	<.001***
B7	64	0.00	0.00	.00
C2	40	0.075	1.78	.08
Total	3062	0.15	23.41	<.001***

*:p<.05, **:p<.01, ***:p<.001

Table 4.20: Results Overreaction bias vs f_b

Planner	# Adj	B^r	t-statistic	p-value
A1	1430	0.13	14.52	<.001***
A3	299	0.13	6.78	<.001***
B1	244	0.21	8.11	<.001***
B3	240	0.18	7.32	<.001***
A2	220	0.19	7.18	<.001***
B2	163	0.20	6.41	<.001***
A4	67	0.06	2.04	.04*
C1	65	0.09	2.55	.01*
B7	64	0.06	2.05	.04*
C2	40	0.15	2.62	.01*
Total	3062	0.15	22.96	<.001***

*:p<.05, **:p<.01, ***:p<.001

Tables 4.21 and 4.22 give an overview of all the biases measured per planner in relation to the statistical forecast and the forecast before. From these overviews it is difficult to draw clear conclusions on the impact of the biases. The idea of this analysis is to show that the biases are present and with the experiment, the goal is to see the effects of the decisional guidance on the presence of the biases. For the overreaction bias there is some clear impact which is mentioned before in this section. For the other two biases it is more difficult to specify on their impacts. What is visible and interesting to see in Table 4.21, that the only two planners (A4 and B7) having a positive $FVA_{s,a}$ also have limited biases. Planner B7 has no biases, while planner A4 only has a significant overreaction bias but an average bias that is far below the average of all adjustments.

Table 4.21: Overview biases per planner vs f_s

Planner	# Adj	B^o	B^a	B^r	$MAPE_a$	$FVA_{s,a}$
A1	1430	0.17***	0.58*	0.14*	47.60%	-5.84%
A3	299	0.09*	0.65*	0.13*	53.12%	-3.58%
B1	244	0.40***	0.47	0.19*	64.89%	-14.81%
B3	240	0.07	0.70*	0.18*	58.43%	-11.00%
A2	220	0.12*	0.61*	0.18*	53.33%	-2.52%
B2	163	0.29***	0.59	0.29*	64.99%	-14.36%
A4	67	0.08	0.63	0.09*	55.19%	4.79%
C1	65	0.14	0.76*	0.20*	46.17%	-0.60%
B7	64	-0.44***	0.00	0.00	58.34%	0.35%
C2	40	0.30*	0.60	0.08	61.69%	-21.22%
Total	3062	0.16***	0.61***	0.15***	52.92 %	-6.72 %

*:p<.05, **:p<.01, ***:p<.001

Table 4.22: Overview biases per planner vs f_{b}

Planner	# Adj	$\mathrm{B^o}$	$\mathrm{B^a}$	$\mathrm{B^r}$	$MAPE_a$	$\mathrm{FVA_{b,a}}$
A1	1430	0.17***	0.60*	0.13*	47.60%	1.08%
A3	299	0.09*	0.70*	0.13*	53.12%	-3.19%
B1	244	0.40***	0.47	0.21*	64.89%	0.20%
B3	240	0.07	0.64*	0.18*	58.43%	1.24%
A2	220	0.12*	0.61*	0.19*	53.33%	0.05%
B2	163	0.29***	0.61*	0.20*	64.99%	27.88%
A4	67	0.08	0.64	0.06*	55.19%	8.12%
C1	65	0.14	0.82*	0.09*	46.17%	3.60%
B7	64	-0.44***	0.83*	0.06*	58.34%	5.09%
C2	40	0.30*	0.63	0.15*	61.69%	-16.86%
Total	3062	0.16***	0.62***	0.15***	52.92 %	1.96%

*:p<.05, **:p<.01, ***:p<.001

4.7 Key findings

The results of the data analysis will form the start of the experiment. The results will be used as input for the experiment to know which behavior planners show in the adjustment process and what this behavior means for the performance of the forecast process. From these conclusions, the focus of the experiment can be clarified.

One of the most important conclusions is that the adjustments currently deteriorate the forecasts' performance. The FVA_s is -6.72 %, meaning that the forecasts have a bigger error after the adjustments than the statistical forecasts. When analyzing the exact behavior of planners to find out the reasons for this poor performance, this leads to multiple findings. First, in general, planners are quite good at determining the direction of an adjustment. However, not all of the adjustments in the right direction lead to an improved forecast. When overreacting occurs, planners deteriorate the forecast by overshooting the actual demand while correcting the adjustment in the right direction. This shows it can be very valuable to give more guidance to planners in determining the size of the adjustment to reduce the overreaction bias. Comparing the FVA of adjustments where the right direction was chosen with adjustments from the wrong direction, it is visible that planners are quite good in choosing the right direction but do not get much value out of these adjustments. This highlights the possibility of using decisional guidance to improve the planner's decision on the size of the adjustment.

When focusing on the performance in the different segmentation categories, the differences for the FVA scores in the ABC categories are quite the same, meaning no significant differences were found. For XYZ categories, the $FVA_{s,a}$ and $FVA_{b,a}$ were significantly higher in Z categories compared to X and Y. When looking at the $MAPE_a$, it is visible that this result is mainly due to the performance of the statistical forecasts, since there are no significant differences in the MAPE scores comparing category Z with X and Y. The $MAPE_s$ in categories X and Y is way lower compared to category Z, meaning the statistical forecasts are already performing quite well in these categories. This results in the fact that planners do add value in category Z but deteriorate the forecasts in categories X and Y. When converting this to providing appropriate guidance when adjusting category X and category Y items, it seems to be logical to also focus on guidance in making a decision whether an adjustment is necessary or not besides focusing on the size of the adjustment. For items of category Z, the focus will only be on guidance for the decision on the size of the adjustment since planners are better at making the adjustment and direction decisions.

Regarding the biases, the most important conclusion is that cognitive biases are present at company A based on the adjustments. However, for the anchoring and optimism biases, it is hard to elaborate in detail on the impact of the biases. For the overreaction bias, it is clear that overreacting results in a negative FVA. Therefore, this will be a focus for the experiment to see whether decisional guidance can reduce the overreaction bias in order to improve the performance of the adjustments. For all the biases, it will be monitored if the decisional guidance impacts the presence of the biases.

Chapter 5

Experiment

In this chapter the hypotheses and the experimental set-up are explained.

5.1 Hypotheses

Based on the results of Chapter 4, the focus is on the effect decisional guidance can have in order to improve forecast accuracy. This is done in an experiment that is further explained in Section 5.1. The most important goal is to see if decisional guidance does add value to the performance of the forecasts. Therefore, all adjustments after and before receiving guidance are compared with each other to test the effect of decisional guidance on the MAPE measured by the FVA in judgmental forecasting. Since many studies found the effectiveness of decisional guidance on the performance of tasks (Montazemi et al., 1996; Fildes et al., 2006; Silver, 1991; Parikh et al., 2001), the same expectation holds for the results of this experiment resulting in the hypothesis below.

Hypothesis 1: Decisional guidance on judgmentally adjusting forecasts has a positive effect on forecast performance.

Performance is measured by the difference in MAPE after and before receiving guidance, which is indicated by the $\text{FVA}_{\text{b,a}}$. Next to general performance, the influence of decisional guidance on task complexity is also tested. The XYZ classification will represent text complexity. The XYZ classification ranks items according to the level of uncertainty of demand determined by the coefficient of variation (Scholz-Reiter et al., 2012). This means that forecasting items of category Z can be seen as more difficult compared to items of category X or Y. Since Montazemi et al. (1996) found in their research that when using decisional guidance, suggestive guidance works better for the less complex task compared to informative guidance. Informative guidance works better than suggestive guidance for more complex tasks (Montazemi et al., 1996). Combining these findings from the literature leads to the following two hypotheses.

Hypothesis 2: Suggestive guidance outperforms informative guidance when adjusting items classified as X or Y.

Hypothesis 3: Informative guidance outperforms suggestive guidance when adjusting items classified as Z.

To be able to test this properly, this means there are four different combinations necessary to test these hypotheses: suggestive guidance on an X or Y item, suggestive guidance on a Z item, informative guidance on an X or Y item, and suggestive guidance on a Z item.

The results from Chapter 4 also show that planners are better in adjusting forecasts of items of category Z compared to items from categories X and Y. This is mainly caused by the fact that the statistical forecasts of items from category Y and especially X perform significantly better compared to statistical forecast from category Z. For this reason, when adjusting items from category X or Y planners also receive guidance on whether an adjustment is necessary or not. This is included in the experiment to test whether such guidance works better than guidance on the size for these types of adjustments. According to Fildes et al. (2006) absolute restrictiveness can be very dangerous since it might be frustrating to the users, and it can be difficult for the designer to determine which processes can be the most relevant to use. However, this does not mean restrictiveness is a bad option as decisional guidance in all cases. Subtle restrictiveness in low-effort tasks can lead to an increase in the accuracy of judgmental forecasts (Fildes et al., 2006). These findings from the literature in combination with the results, lead to the idea of testing different types of guidance for low-effort tasks such as adjusting X and Y items. Next to the decisional guidance on the size of the adjustment, in the experiment, participants will receive guidance on the decision if an adjustment is necessary. This will be compared with the guidance on the size, and the expectation is that the decision on whether the adjustment is necessary will be more beneficial since this can be seen as subtle restrictiveness on low-effort tasks.

Hypothesis 4: Decisional guidance on deciding if an adjustment is necessary outperforms decisional guidance on the size of the adjustment in categories X and Y.

To test these hypotheses properly, there are three different combinations necessary to test these hypotheses: suggestive guidance on the size of the adjustment, suggestive guidance on the necessity of the adjustment, and informative guidance on the size of the adjustment. The differences will be tested in an experiment that is set up such that it looks similar to the real-life situation from which the data of Chapter 4 comes. A detailed explanation of the exact approach follows in Section 5.1.

Table 5.1: Overview hypotheses

Hypothesis
1. Decisional guidance on judgmentally adjusting forecasts has a positive effect on the forecast performance.
2. Suggestive guidance outperforms informative guidance when adjusting items classified as X or Y.
3. Informative guidance outperforms suggestive guidance when adjusting items classified as Z.
4. Decisional guidance on deciding if an adjustment is necessary outperforms decisional guidance on the size of the adjustment in category X and Y.

5.2 Experimental set-up

In order to be able to test the hypotheses, an experiment is set up to test decisional guidance. The experiment is set up in Qualtrics. First, a brief introduction of the experiment is given, followed by a more detailed explanation of the participants' task. The main goal of the experiment is to test the effects of decisional guidance on the performance of the adjustments of planners on forecasts. As stated in Section 5.1, the goal of this experiment is to test whether decisional guidance can improve the forecast accuracy (measured with the MAPE and FVA) and help planners make better adjustments. During the experiment, a situation is created that is as close as possible to the real-life situation of demand planners. As participants, employees from EyeOn are invited to participate in the experiment. Therefore, all participants have enough basic knowledge about forecasting, which is necessary to be a representative participants in this experiment. They have all relevant experience in the supply chain world and are, therefore, a good representative group of participants.

5.3 Judgmental forecasting task

The experiment consists of an introduction section with all relevant information on the task. The task and role of the participant during the experiment are explained during the introduction phase. After the introduction, the participant gets three training tasks to get familiar with the tasks. During the experiment, the participants acted as a planner and got the task of judging whether the statistical forecast needed to be adjusted. In real life, situation planners have a lot of information to use in their decision-making process. Planners can see historical data on statistical forecasts, final forecasts and the actual demand for items. Furthermore, planners have much more information from internal and external sources that are not included in the statistical forecast and might be a reason for them to adjust the forecast. In the experiment, this situation is simulated as realistically as possible. Therefore, the participants receive historical data on the item they need to forecast. The historical information of one item stays the same during the entire experiment to make sure the participant has the same information for each task.

5.3.1 Details tasks

The items that the participants need to forecast are ice creams, and the features are price and temperature. Ice creams are known items, and the impact of price and temperature on this is easy to imagine. It is chosen to include as extra, more detailed information for planners to review the statistical forecast. All three of them are chosen to make sure the participants will understand the task at hand. In the experiment, both the price and temperature can fluctuate in an upwards and downward direction for every period. For the price, downward changes may be more realistic (promotions), but this choice has been made not to steer the participant in the direction of an increase in the forecasted amount. The historical information provides information on three previous periods with the statistical forecast, the actual temperature, the actual price and the actual demand of those periods. The statistical forecast is based on weighted averages of those two features, which is also indicated to the participants. The features act as the latest accurate information that planners have in real-life situations that should give the participants a reason to review the statistical forecast and make a judgment on whether the forecast needs to be adjusted.

There are three different items on which the participants need to review and decide if an adjustment is necessary during various periods. There is one X-item, one Y-item and one Z-item. To make it easy to recognize, flavors of ice creams are used to indicate different products to forecast.

All input values for the experiment can be found below in Table 5.3. The numbers of the statistical forecasts and the features are randomly generated for each item within a certain range with a uniform distribution. The range of the numbers of the statistical forecast differs per item. For the features, the range is the same for each item. The formula for the actual demand is visible below. The demand is based on the statistical forecast since the statistical forecast is based on the average temperature and price in the experiment. The noise is also randomly generated (uniform distribution) for each item and has a different range for X, Y and Z-items that can be found below in Table 5.3.

$$\text{Actual demand:} s^{\mathrm{i}} = f_{\mathrm{s}} + \beta_{\mathrm{p}}\Delta_{\mathrm{p}} + \beta_{\mathrm{t}}\Delta_{\mathrm{t}} + \epsilon \tag{5.1}$$

Table 5.2: Explanation variables

Variable	Definition	Formula if applicable
s^{i}	Actual demand in period i	
$\mathrm{f_s}$	Statistical forecast	
β_{p}	Weighting factor of price difference effect	
β_{t}	Weighting factor of temperature difference effect	
p_{a}	Average price	
p_{i}	Price of product in period i	
t_{a}	Average temperature	
t_{i}	Temperature of product in period i	
Δ_{p}	Price difference	$p_{\mathrm{a}} - p_{\mathrm{i}}$
Δ_{t}	Temperature difference	$t_{\mathrm{a}} - t_{\mathrm{i}}$
ϵ	Noise	

In Table 5.3 the differences between the ranges of the variables used to generate the demand are shown. The base price and range are chosen to keep it as understandable as possible for the participants. The same holds for the temperature and to keep it also quite realistic. Weighting factors are chosen to keep a similar ratio between the price and temperature of all three items. This has been done such that it does not get too complicated to understand the effects of the price and temperature differences. The difference between the weighting factors has been chosen such that the difference between the statistical forecast error and the coefficient of variation (used for classification XYZ) is as realistic as possible. The noise term is different among the different items to make sure the statistical errors are different also when actual price and temperature would be equal to the averages. These input values cause also a difference in the quality of the decisional guidance among the X, Y and Z-items. However, it is still interesting to see the effect of the decisional guidance among the different items. Furthermore, the different ranges for the noise ensure that it is more difficult for the system to compute the guidance for Z-items compared to Y and X-items.

Table 5.3: Input values experiment

Metric	X	Y	Z
β_p	5	40	50
β_t	0.5	4	5
p_a	5,00	5,00	5,00
t_a	20	20	20
Range p_i	$[4.00, 6.00]$	$[4.00, 6.00]$	$[4.00, 6.00]$
Range t_i	$[10, 30]$	$[10, 30]$	$[10, 30]$
Range ϵ	$[-5, 5]$	$[-10, 10]$	$[-15, 15]$

In Table 5.4, the statistical forecast errors and the coefficient of variation are shown. The numbers of the coefficient of variation are not completely in line with the thresholds used at EyeOn, but they show significant differences between the demand patterns of the items, which is the main goal. Furthermore, if the coefficient of variation was the same as the thresholds of EyeOn, this led to very big differences and an unrealistic demand pattern that might influence the results of the experiment.

Table 5.4: Statistical error and coefficient of variation items

Variable	X	Y	Z
CV	0.18	0.46	0.67
MAPE_s	4.48 %	52.40%	118 %

Table 5.5: Overview differences in tasks per hypothesis

	Differences in tasks		
Hypothesis	Forms of guidance	Items	Type of guidance
H1	Guidance vs No Guidance		
H2	Informative vs Suggestive	X,Y vs Z	
H3	Informative vs Suggestive	X,Y vs Z	
H4	Informative vs Suggestive	X,Y	On adjustment vs on Size

An example of a basic forecasting task is visible below in Figure 5.1.

Training task 2

You will now receive some historical information about the item you are going to forecast.
Remember the base price is €5.00 and the average temperature 20°C, which are both used for the statistical forecast.

Product ID	Item classification
Salted Caramel	Y

Historical information				
Period	Forecast	Price	Temperature	Demand
1	78	€5,50	16°C	43
2	55	€4,70	25°C	86
3	50	€6,00	26°C	26

Below you see the specific information about the period you are going to forecast.
Keep in mind that the statistical forecast also has a certain error.
You are expected to review the forecast of period 6 and decide if an adjustment is necesarry.
Please give the final forecast in the text box below.

Forecast				
Period	Forecast	Price	Temperature	Demand
6	63	€5,40	19°C	?

Figure 5.1: Example task experiment

5.4 Decisional guidance during the experiment

When the participant has finished the training tasks, the participant receives nine tasks of one of the three forms of guidance as described earlier, such that for each item, there are three tasks of guidance. The order of the three forms of guidance is randomized. In total, the participant receives 27 tasks with decisional guidance. We use a 3x3 factorial design, i.e. each participant needs to make decisions on each of the three products and receives all three forms of guidance. This results in nine unique tasks that all have three repetitions with different data but the same design(guidance and item category). The three repetitions are chosen in order to receive enough data to analyze the data without doing the experiment too long. The three different forms are suggestive guidance on the size of the adjustment, informative guidance on the size and suggestive guidance on the adjustment itself. Examples of the three possible forms of guidance are provided below in Figures 5.2, 5.3, and 5.4. In the figures, only the guidance part is visible. In the experiment, the participants also saw the historical and actual information on their screens when they received the guidance. Suggestive guidance on the size (SG) provides a suggestion (a value) for the adjustment, while informative guidance (IG) on the size provides information like past errors. Suggestive guidance on the adjustment (SGA) provides the suggestion of keeping the forecast at the same number as the statistical forecast. For the guidance on the size of the adjustments, the guidance is calculated based on the formula of the demand without the noise (ϵ). This situation with the guidance having the knowledge about the β_p and β_t is different from a situation that would appear in a real-life situation. However, it could represent a situation of a real-life situation in which there is a lot of information on the demand based on specific events. Furthermore, the scope of this master thesis is not to research the best possible way to create or calculate decisional guidance. Suggestive guidance really provides the number that the system suggests based on the earlier information and informative guidance provides a percentage (corresponding to that number) calculated by the system. Both the suggestive and informative guidance have, therefore, the same level of quality of guidance but differ in the tone and the way of delivering the information.

If the initial final forecast that the participant submits is the same as the guidance, the guidance will not be provided to the participant. As an example, if the suggestive guidance suggests 53 and the answer of the participant is also 53 for the regular task, the participant will continue with the next task. This holds for all types of guidance, also for percentages. This causes the fact that the number of tasks with guidance does not have to equal the number of tasks without guidance a participant received.

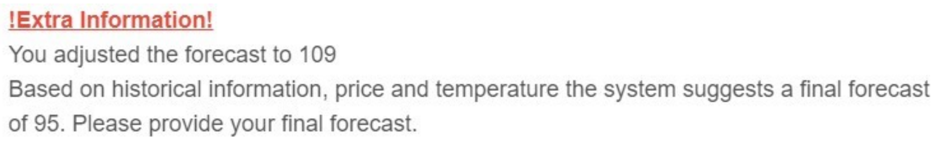

Figure 5.2: Example suggestive guidance

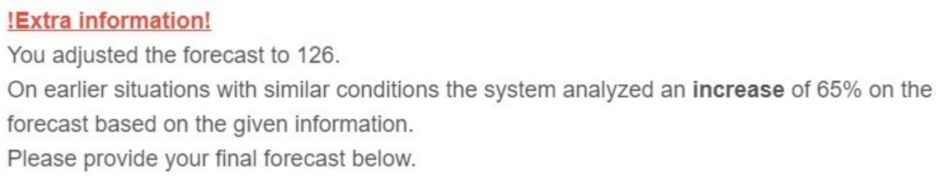

Figure 5.3: Example informative guidance

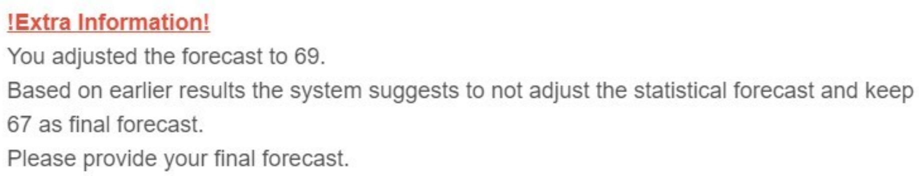

Figure 5.4: Example suggestive guidance on the adjustment

Chapter 6

Results experiment

In this section, the results of the experiment are shown, and the conclusions have been drawn based on this. First, the results of the hypotheses of Section 5.1 are shown and discussed. The hypotheses are tested with a t-test and a Wilcoxon test. Based on whether it is appropriate for the dataset used, a paired or unpaired test is used. Paired tests are used in Section 6.2 since, in this case, a comparison can be made with MAPE before the guidance and after the guidance to see the actual effect. For all other tests in which the comparison is made between categories or the form of guidance unpaired tests are used to check for statistical difference between the $FVA_{b,a}$. An explanation of the statistical tests can be found in Appendix A. Afterward, the data have been further analyzed for exploratory analysis regarding the cognitive biases and the behavior of the participants on the different forms of guidance.

In this analysis, multiple metrics have been used to analyze the data. The main metrics are similar to the metrics used in Chapter 4. The metrics used are the same, but in the experiment, there were two forecasts given by the participant, one before the guidance and one after receiving the guidance. Compared to the data of Chapter 4, the forecast before is, in this case, the forecast given by the planner before receiving guidance, and the forecast after is the adjusted forecast of the planner after receiving guidance. In case the planner did not receive decisional guidance, the forecast before equals the forecast after.

6.1 General statistics

In total, there were 36 participants who fulfilled the experiment, which led to a total of 972 different forecasting tasks in which the participant received 837 times an extra notification with decisional guidance. This means that 135 times (13,89%), the initial adjusted forecast of the participant was equal to the decisional guidance, and therefore, they did not receive the decisional guidance. Out of the 135 observations where no decisional guidance was given, the majority was in part of SGA (85,19% versus 5.19% SG and 9.63% IG).

Table 6.1: General statistics experiment

What?	# of observations
Participants	36
Tasks without guidance	972
Tasks with guidance	837
Adjusted statistical forecast	672
Adjusted adjusted forecast	462

Regarding the performance of the participants in adjusting the statistical forecasts, a similar pattern was found compared to the results in Chapter 4. On average, the participants perform well in deciding on the direction of the adjustment. However, participants did not always improve the forecast when adjusting in the right direction. This was especially visible for X-items, wherein the majority of the observations, an adjustment in the right direction did not lead to an improvement. In 74.64% of the observations, the direction of the initial adjustment was correct, but only 20.75% of these observations led to an improvement in the forecast accuracy. For Y-items and Z-items, a majority of the adjustments in the right direction did improve the statistical forecast, but still, some observations that were in the right direction decreased the forecast accuracy.

Table 6.2: Right direction initial adjustment vs f_s

Category	N	Right direction	Wrong direction
X	213	74.65%	25.35%
Y	242	81.40%	18.60%
Z	217	69.59%	30.41%
Total	672	75.45%	24.55%

Table 6.3: Improved initial adjustment vs f_s

Category	N	Improved	Same	Decreased
X	213	15.49%	2.82%	81.69%
Y	242	69.42%	1.24%	29.34%
Z	217	60.83%	1.38%	37.79%
Total	672	49.55%	1.79%	48.66%

In Table 6.4 the general results of the experiment are visible. This showed that initially the participants on average did not add value to the statistical forecasts. After receiving the guidance the planners did add value to the statistical forecast and also had on average a lower error compared to the statistical forecast. This implies that providing decisional guidance is beneficial.

Table 6.4: General results experiment

Metric	Overall result
$MAPE_s$	43.85%
$MAPE_b$	44.79%
$MAPE_a$	34.72%
$FVA_{s,b}$	-0.94%
$FVA_{s,a}$	9.13%
$FVA_{b,a}$	10.07%

6.2 Effect decisional guidance

From Table 6.4 can already be concluded that the presence of decisional guidance led to a decrease in the forecast error. To test the effect of providing decisional guidance to the participants in the experiment properly, the regular tasks without guidance are compared to the tasks with guidance. In case a participant did not receive any decisional guidance, this observation is not taken into account for this analysis. There have been multiple tests performed, with different datasets. These results are visible in Table 6.5.

Table 6.5: Results test effect decisional guidance

Type of guidance	MAPE		T-test		M-W test	
	Guidance	No guidance	t-statistic	p-value	Z-score	p-value
All	34.71%	44.79%	9.128	<.001**	-12.391	<.001**
SG	37.73%	52.07%	6.861	<.001**	-9.790	<.001***
IG	28.76%	38.61%	6.369	<.001***	-7.541	<.001***
SGA	39.01%	42.95%	1.987	.048*	-2.656	.008**

*:p<.05, **:p<.01, ***:p<.001

The results from Table 6.5 show that providing decisional guidance when judgmentally adjusting forecasts lead to a significantly lower error and therefore, value was added to the forecasts by the decisional guidance. Based on these results, Hypothesis 1 is confirmed. In addition to that, also each form of guidance has a statistically significant effect on the forecast performance measured by the MAPE. With the knowledge that all forms of guidance individually have a significant effect on the forecast performance, in the next section, we performed analyses to check if there are significant differences between the best form of guidance for each category.

6.2.1 Suggestive versus informative guidance

With the knowledge that all forms of guidance have a significant effect on forecast performance, more analysis has been done on the effect of the different forms of decisional guidance. In the experiment, there were two different forms of guidance on the size of the adjustment and one form on the adjustment itself. Since these forms differ a lot in this analysis, the focus was on the differences between suggestive versus informative guidance on size. Table 6.6 shows the overall effects of all forms of guidance. The big difference between the $MAPE_s$ among SG versus IG and SGA originates from category

Z, in which one observation has a very low actual demand value and a high statistical forecast, resulting in a very high MAPE.

Table 6.6: Performance guidance forms

Metric	IG	SG	SGA
$MAPE_s$	35.18 %	57.31 %	36.32 %
$MAPE_b$	38.61 %	52.07 %	42.95 %
$MAPE_a$	28.76%	37.73%	39.01 %
$FVA_{s,b}$	-3.43%	5.24 %	-6.63 %
$FVA_{s,a}$	6.42%	19.58 %	-2.69 %
$FVA_{b,a}$	9.85%	14.34%	3.93 %

Table 6.6 shows some differences between the effect of suggestive and informative guidance in general, but no statistically significant differences were found. In line with the hypotheses, the differences between the effectiveness of suggestive guidance and informative guidance among the different categories have been tested in the results are shown in the next section.

6.3 Categories

In the experiment, there were three different items that each had a different categorization according to the XYZ analysis. Since these items are ranked based on their stability of demand, this is also related to how predictable these items are. In Chapter 4, it became clear that planners add more value to the forecasts with their judgmental adjustments when the demand is less stable, especially more for Z items and to a less extent for Y items. Similar findings are found in the experiment. For X-items, the participants did not add value to the statistical forecast, but for Z and Y-items, the participants added value to the statistical forecast in terms of accuracy. The difference between Z and Y items is different compared to the earlier results. While in Chapter 4, planners only added value in category Z-items, the participants of the experiment also added value for the Y-items.

Table 6.7: Performance among XYZ-items

Metric	X	Y	Z
$MAPE_s$	4.42 %	45.07 %	80.87 %
$MAPE_b$	16.98 %	31.89 %	82.57 %
$MAPE_a$	14.09%	23.48%	67.85 %
$FVA_{s,b}$	-12.56%	13.19 %	-1.70 %
$FVA_{s,a}$	-9.67%	21.60 %	13.02 %
$FVA_{b,a}$	2.89%	8.41%	14.72 %

Table 6.8: Performance different forms among X-items

Metric	SG	IG	SGA
$MAPE_s$	3.56 %	5.09 %	4.52 %
$MAPE_b$	22.92 %	17.32 %	16.58 %
$MAPE_a$	18.27%	15.10%	13.26 %
$FVA_{s,b}$	-19.36%	-12.23 %	-12.06 %
$FVA_{s,a}$	-14.71%	-10.01 %	-8.74 %
$FVA_{b,a}$	4.65%	2.22%	3.32 %

Table 6.9: Performance different forms among Y-items

Metric	SG	IG	SGA
$MAPE_s$	53.90 %	49.62 %	36.88 %
$MAPE_b$	28.68 %	40.06 %	30.69 %
$MAPE_a$	17.05%	26.00%	30.88%
$FVA_{s,b}$	25.22 %	9.56 %	6.18 %
$FVA_{s,a}$	36.85%	23.62 %	6.00 %
$FVA_{b,a}$	11.63%	14.06%	-0.18 %

Table 6.10: Performance different forms among XY-items

Metric	SG	IG	SGA
$MAPE_s$	28.85 %	28.10 %	21.69 %
$MAPE_b$	25.82 %	29.07 %	24.07 %
$MAPE_a$	17.66%	20.73%	22.61 %
$FVA_{s,b}$	3.04%	-0.97 %	-2.38 %
$FVA_{s,a}$	11.19%	7.37 %	-0.92 %
$FVA_{b,a}$	8.16%	8.34%	1.46 %

6.3.1 X and Y-items

In Table 6.10 it is already visible that there is a higher number for forecast value add by the decisional guidance when this guidance is informative compared to suggestive guidance. This would be in line with Hypothesis 2. However, to test this properly a student t-test and a Mann Whitney U test have been performed. These tests do not show any statistical significant differences between suggestive guidance and informative guidance which is visible in Table 6.12. When looking at the results of the forecasts of item X and Y separately, there are some differences visible but none of them were statistically significant which can be found in Table 6.11.

Table 6.11: Results statistical tests different forms among X and Y-items

Category	Metric	Guidance type		T-test		M-W test	
		SG	IG	t-statistic	p-value	U-score	p-value
X	$FVA_{b,a}$	4.65%	2.22%	1.587	.114	4882.00	.358
Y	$FVA_{b,a}$	11.63%	14.06%	-0.810	.210	5550.00	.784

*:p<.05, **:p<.01, ***:p<.001

Table 6.12: Results statistical tests different forms among XY-items combined

Guidance type		Metric	Guidance type		T-test		M-W test	
1	2		1	2	t-statistic	p-value	U-score	p-value
IG&SG	SGA	$FVA_{b,a}$	8.25%	1.46%	4.261	<.001***	23659.00	<.001***
IG	SGA	$FVA_{b,a}$	8.34%	1.46%	3.615	<.001***	12121.00	<.001***
SG	SGA	$FVA_{b,a}$	8.16%	1.46%	4.569	<.001***	11538.00	<.001***
IG	SG	$FVA_{b,a}$	8.34%	8.16%	-0.104	.918	21334.00	.670

*:p<.05, **:p<.01, ***:p<.001

6.3.2 Z-items

In Tables 6.13 and 6.14 the results of the tasks in Category Z are visible. From these tables, the forecast performance after receiving informative guidance is significantly better compared to the performance after receiving suggestive guidance. This is measured by the MAPE. However, looking only at this measurement method would give an incomplete overview of the situation, as stated before. It is more interesting to look at the effect that is caused by the decisional guidance. Therefore, the $FVA_{b,a}$ is used to analyze this effect. The $FVA_{b,a}$ for suggestive guidance is higher than the $FVA_{b,a}$ for informative guidance, but the difference is not significantly different according to the performed t-test and Mann-Whitney test, which is visible in Table 6.14. This result is contradictory to what was expected according to the literature and stated in hypothesis 3.

Table 6.13: Performance different forms forms among Z-items

Metric	SG	IG	SGA
$MAPE_s$	112.38 %	49.70 %	71.01 %
$MAPE_b$	102.89 %	58.15 %	87.71 %
$MAPE_a$	76.57%	45.21%	77.92%
$FVA_{s,b}$	9.49 %	-8.45 %	-16.70 %
$FVA_{s,a}$	35.81%	4.49%	-6.91 %
$FVA_{b,a}$	26.32%	12.94%	9.79 %

Table 6.14: Results statistical tests different forms Z-items

Metric	Guidance type		T-test		M-W test	
	SG	IG	t-statistic	p-value	U-score	p-value
$FVA_{b,a}$	26.32%	12.94%	1.959	.052	4730.00	.069

*:p<.05, **:p<.01, ***:p<.001

6.4 Effect of suggestive guidance on adjustment

The last hypothesis stated that suggestive guidance on the adjustment for X and Y items would result in higher performance of the $FVA_{b,a}$ than decisional guidance on the size of an adjustment. The idea behind the hypothesis was that it is better not to adjust items in categories X and Y at all because of a stable demand pattern and the results of Chapter 4 that showed a negative impact of the adjustments on the forecast performance.

When looking at all data from categories X and Y combined, SGA guidance has a lower $FVA_{b,a}$ compared to SG and IG. Also, the statistical tests performed indicate a significant difference, which is shown in Table 6.12. The reason for these results is the difference in the performance in Y categories in the experiment compared to the results of Chapter 4. In the experiment Table 6.11, shows that the participants clearly added value to the forecasts for Y-items while they did not do that in Chapter 4. Therefore, to get a better understanding of the effect of SGA, analyses are also performed only on category X-items.

The general results in Table 6.15 show a lower average MAPE for the tasks that received suggestive guidance on the adjustment. However, comparing the added value to the forecasts before and after receiving guidance, it does not show a clear difference in performance.

The reasoning for why this type of advice does not work could be the quality of the advice. To understand the working of this advice, the advice should be divided into two parts. One part of the guidance is the quality of the guidance, and the other part is the fact that people change their behavior and listen to the guidance. In this case, it seems that the participants did listen to their behavior. When the judgmental forecast is equal to the number provided by the guidance, the participant does not receive the guidance anymore, and his initial adjusted forecast will be the final forecast. For the suggestive guidance on adjustment, the guidance gives participants the advice to not adjust the statistical forecast and keep the statistical forecast as the final forecast. Therefore, the goal of this guidance is to convince participants not to adjust the statistical forecast. When analyzing the results, the $FVA_{b,a}$ is not higher for the tasks with SGA compared to SG or IG. However, the final error is the lowest for the tasks with the SGA. The student t-test showed a statistically significant difference between the $FVA_{s,a}$ of SGA versus IG and SG combined. However, the Mann-Whitney test did not show statistical significance, which means the statistical significance can not be proven, which is visible in Table 6.16. A reason for the fact that the result of a lower $FVA_{s,a}$ can be important to analyze this situation is because of the setup of the experiment. Participants do not always adjust the statistical forecast at all, especially for X items. For X-items, this happens in 34.26% of the cases, as can be seen in Table 6.17. This group of people does not receive the SGA at all since the goal of this guidance is to not adjust, and they have already achieved that goal. So when looking at the total number that, in the end, has exactly the number of the guidance as the final forecast, this is way higher for SGA compared to SG and IG. Therefore, this type of guidance still could add value to the forecast accuracy, although the difference could not be statistically proven to be better than SG and IG.

Table 6.15: Performance different forms among X-items

Metric	SG	IG	SGA
$MAPE_s$	3.56 %	5.09 %	4.52 %
$MAPE_b$	22.92 %	17.32 %	16.58 %
$MAPE_a$	18.27%	15.10%	13.26 %
$FVA_{s,b}$	-19.36%	-12.23 %	-12.06 %
$FVA_{s,a}$	-14.71%	-10.01 %	-8.74 %
$FVA_{b,a}$	4.65%	2.22%	3.32 %

Table 6.16: Results statistical tests different forms among X-items

Metric	Guidance type		T-test		M-W test	
	IG & SG	SGA	t-statistic	p-value	U-score	p-value
$FVA_{s,a}$	-11.71%	-8.74%	-3.243	<.001***	10500.00	.136

*:p<.05, **:p<.01, ***:p<.001

Table 6.17: Guidance as final forecast X-items

Guidance	% Exactly taken guidance
SG	27.78%
IG	30.56%
SGA	51.85%

Based on the results of the hypotheses no clear differences were found between SG and IG among the different categories X,Y and Z. However, for SGA it became clear that only for category X items SGA could be of value since planners do not add value there in general with their judgmentally adjustments.

6.5 When to use guidance on size

In the previous section, the hypotheses have been handled and focused on the differences between the forms of guidance. For the guidance on the size of the adjustment, it is interesting to see if there are specific situations in which the guidance has a positive effect and when it does not have a positive effect. The first thing that becomes clear from analyzing the data is the difference between participants in willingness to accept the guidance and listen to the guidance. On average, in 53% of the cases, a participant received guidance the participant "listened to guidance". Listening to guidance is achieved when a participant changes his forecast after receiving the guidance in the direction of the guidance. However, this percentage of listening to guidance differs a lot per participant ranging from 0 percent to 100%. This indicates it is important to keep track of an individual's behavior regarding the guidance since this might deviate a lot among a group of individuals.

In total, there are 63 observations in which the $\mathrm{FVA}_{\mathrm{b,a}}$ is negative. In the majority of these cases (57.14%), the adjusted forecast already improved the statistical forecast. This means it would be better not to have given guidance to the participants. In those cases, the adjustment itself already improved the statistical forecast and therefore, the guidance that increased the error can be seen as harmful. Although the absolute numbers are not very big, it is an interesting finding that the decisional guidance in some cases deteriorated the forecast accuracy when providing the guidance was harmful since the performance was already improved by the adjusted forecast.

Whether decisional guidance is effective depends on the quality of the decisional guidance and the willingness of the participants to change their behavior based on the decisional guidance. The willingness to change is measured with "listening to guidance" as mentioned above. Furthermore, a closer look at the percentage of the guidance that has been fulfilled by the participant can be interesting. When the decisional guidance suggests a final forecast of 36 with the adjusted forecast being 32, a final forecast of 33 by the participant means the participant fulfilled 25% of the decisional guidance.

Table 6.18 shows the difference between accepting the guidance that is in the same direction or the opposite direction compared to the direction suggested by the participant. It shows the situation a higher percentage for listening to the guidance when it is in the same direction. This could imply that the participants are more willing to accept decisional guidance that is in the same direction as their own adjustment.

Table 6.18: Percentage of the decisional guidance that is fulfilled with the adjustment

Direction guidance	Percentage listened to guidance	Percentage of guidance fulfilled
Same direction	63.1%	86.0%
Opposite direction	50.9%	75.0%

Besides the earlier mentioned direction of the guidance, the size of the guidance also could impact the willingness of participants to accept the decisional guidance. A closer analysis showed that there is little to no difference in the listening to guidance among the size of the guidance. However, some interesting findings are found when looking at a fraction of the guidance that is fulfilled. This was slightly higher for IG compared to SG(64% versus 50%). This difference was mainly caused by the observations that received big advice from the decisional guidance. Small advice was defined as small when being smaller than the average advice size. The same holds for big adjustments in the opposite direction (different direction of the advice compared to the initial adjustment). Overall, small adjustments have a higher percentage of being fulfilled compared to big adjustments (72% versus 40%). For both IG (83% for big versus 39% for small) and SG (60% for big versus 41% for small), this difference was visible, but for IG, the percentage that was fulfilled was way higher for big advice.

Another point to take into account when analyzing the size of the decisional guidance provided is the average $FVA_{b,a}$ of small adjustments versus big adjustments. For small adjustments the average $FVA_{b,a}$ was 3.06 % and for big adjustments the average $FVA_{b,a}$ was 22.26 %. Small adjustments were identified as small when being smaller than average and big adjustments when being bigger than average. Of course, it can be seen as logical that for bigger guidance numbers, the $FVA_{b,a}$ is bigger, but it also raises the question of whether small adjustments are very useful.

Table 6.19: Results hypotheses

Hypothesis	Outcome
1. Decisional guidance on judgmentally adjusting forecasts has a positive effect on the forecast performance.	Supported
2. Suggestive guidance outperforms informative guidance when adjusting items classified as X or Y.	Not supported
3. Informative guidance outperforms suggestive guidance when adjusting items classified as Z.	Not supported
4. Decisional guidance on deciding if an adjustment is necessary outperforms decisional guidance on the size of the adjustment in category X and Y.	Not supported

6.6 Biases

The cognitive biases that have been measured and analyzed in Chapter 4 will also be used in this section. In Table 6.21, all the biases are shown with situations before and after receiving guidance. For the different guidance forms, only the observations where the guidance was actually provided are taken into account. On average, providing guidance on the size results in a slight decrease in biases. To get a good understanding of the effect of the guidance on the individual biases, it is important to understand what the biases imply in relation to the forecast performance. The measurement for optimism bias is related to the average error. Overreaction bias takes the sum of all cases in which the error increased, but the direction was correct. The last bias, anchoring bias, takes the fraction of all adjustments that improved the accuracy but stayed too close to the statistical forecast as anchor value. The results in Table 6.21 are quite similar to the average biases found in Section 6.21.

Table 6.20: Overview biases

		Before guidance			After guidance		
Guidance form	# Adj	B^o	B^a	B^r	B^o	B^a	B^r
All	972	0.20	0.69	0.17	0.16	0.67	0.15
IG	311	0.16	0.84	0.13	0.09	0.73	0.13
SG	317	0.25	0.71	0.21	0.20	0.70	0.20
SGA	209	0.15	0.51	0.26	0.13	0.51	0.26

6.6.1 Biases among different categories

When looking at the biases among the different categories, some clear differences are visible. On average, overreaction is way more present for X-items compared to Y and Z-items. In Table 6.21 can be found the biases before and after receiving decisional guidance. For X-items, it is visible that the decisional guidance manages to some extent to reduce the overreaction bias that was present before receiving the guidance, but there is not a very big difference visible. Furthermore, it is clear overreaction is mainly a problem for X-items and less for Y and Z-items. Another interesting point is the decrease in optimism bias, which is visible for all forms of guidance and in all categories. This could imply that participants, however, being optimistic, are willing to change that behavior when receiving guidance.

Table 6.21: Overview biases

		Before guidance			After guidance		
Guidance form	# Adj	B^o	B^a	B^r	B^o	B^a	B^r
X	274	0.15	0.21	0.43	0.10	0.51	0.40
Y	311	0.03	0.70	0.09	0.03	0.61	0.05
Z	317	0.42	0.83	0.06	0.30	0.82	0.08

Chapter 7

Conclusion

7.1 Data-analysis

The goal of this master thesis was to get more insights into the behavior and performance of planners in judgmentally adjusting forecasts and to test the effects of decisional guidance on the performance of these adjustments. The data analysis on the current behavior and performance of planners showed that, in general, the planners do not add value with their adjustments compared to the statistical forecasts provided by EyeOn. Adjustments were mostly in the right direction, which indicates the potential added value of judgmental adjusting forecasting. However, planners did not add much value with the adjustments that were in the right direction. This was partly due to the overreaction bias, in which planners detect the right direction in which the forecast should be adjusted but overshoot the magnitude of the adjustment and decrease the forecast accuracy. Furthermore, when the direction was chosen correctly, did mostly not lead to very significant improvements in the forecast accuracy. This could be the result of the detected anchoring bias, in which planners adjust in the right direction but focus too much on the "anchor value" and stay too close to the statistical forecast. Both the presence of anchoring and overreaction bias indicate that planners are able to add value to the forecast. Still, the current process is not optimal and can be improved as it deteriorates the forecast accuracy.

To dive deeper into the performance of the planner, a more extensive analysis was performed to detect differences in performance and behavior among the different segmentation categories. Segmentation based on turnover (ABC) did not lead to significant differences. However, segmentation based on stability of demand (XYZ) showed more interesting and significant results. Planners do add value for Z-items but lack in adding value to X and Y-items. This is mainly because of the good performance of the statistical forecast. Especially for X-items and in less extent, for Y-items as well, the performance of the statistical forecast is already very good. This makes it difficult for planners to add value to the forecast, which was also visible in the results. Here we also evaluated the correctness of the direction chosen. We found that planners were better at choosing the right direction for Z-items, which also resulted in adding more value to the forecast of Z-items. From these results could be concluded that for Z-items, it would be very useful to focus on guiding the planner toward the right size of the adjustment, while for X and Y-items, it could also be useful to focus on decisional guidance on the choice to make the adjustment self.

From the data analysis can be concluded that planners do have the ability to detect when adjustments are necessary but have difficulties deciding the right size of an adjustment. This is also partly due to the presence of cognitive biases. The results showed the presence of optimism bias, anchoring and overreaction bias for almost all planners. Especially the overreaction bias has a direct effect on deteriorating the forecast performance. For Z-items, planners do add value to the statistical forecast, while planners do not add value in the current situation for X and Y-items. The reason for this is mainly the fact that the performance of the statistical forecast is way better for X and Y-items compared to Z-items. This leaves less space for planners to improve the forecast performance.

7.2 Effects decisional guidance

With the knowledge about the current performance of the planners, the effects of decisional guidance were tested. We conclude that all three forms (informative guidance on the size, suggestive guidance on the size and suggestive guidance on the adjustment) have a significant effect on the forecast performance by improving the forecast accuracy (Hypothesis 1). In general, suggestive and informative guidance on the size had a bigger impact on the forecast performance than suggestive guidance on the adjustment in terms of $FVA_{b,a}$.

When specifying the analysis on the different categories used for the experiment (XYZ), it became clear that suggestive guidance on the adjustment mainly had a positive effect on X-items. This was partly due to the quality of the decisional guidance on the adjustment compared to guidance on the size. Furthermore, for Z and Y-items, planners are, in most cases, adjusting in the right direction and suggestive guidance on the adjustment takes away the value of that choice of the planner by eliminating a successful adjustment. This is also visible in the $FVA_{b,a}$. For X-items, the suggestive guidance on the adjustment does not show statistically significant differences compared to the guidance on the size, while for Y and Z-items, the guidance on the size outperformed suggestive guidance on the adjustment. This means that for Y and Z-items, decisional guidance on the size works better than guidance on the adjustment. Furthermore, for X-items with the presence of suggestive guidance on the adjustment, the $FVA_{s,a}$ was lower than the $FVA_{s,a}$ for guidance on the size. This could be explained by the fact that not all participants do adjust a statistical forecast, especially for X-items. Once the participant decides not to adjust the statistical forecast, the decisional guidance (SGA) will not be shown. Therefore, in general, more participants will, in the end, have a final forecast in which they did not adjust the statistical forecast compared to when showing the guidance on the size.

Regarding the differences between guidance on the size, the results of the experiment do not show clear differences between the effect of the two different forms of guidance. There were some differences in the performance of the $FVA_{b,a}$, but no significant differences were found (Hypothesis 2 and 3).

An analysis of the cognitive biases showed that in the experiment, cognitive biases (anchoring bias, optimism bias and overreaction bias) were present. Comparing the biases in the situation before receiving the guidance compared to those after receiving the guidance did not led to significant changes in the overall biases. There were some differences visible when comparing the presence among the different items. Overreacting happened more when adjusting X-items, which could be the result of the fact that the performance of the statistical forecast is already quite good for these items. This also highlights why suggestive guidance on the adjustment can be of added value for X-items.

Furthermore, the analysis gave some insights into the effect of decisional guidance on the size of the adjustment in general. When receiving decisional guidance that was in the same direction, this led to a higher percentage of observations in which the participant listened to the guidance. Furthermore, the percentage of decisional guidance that was fulfilled was higher for guidance in the same direction compared to guidance in the opposite direction. This implies that the participants were more intended to accept and listen to decisional guidance if that confirms their choice of direction and suggests overreacting.

The last findings are regarding the size of the decisional guidance. There was no difference found between large and small decisional guidance (the absolute difference between the guidance number and the initial adjusted forecast of the planner) regarding listening to guidance, but the percentage of the guidance that was fulfilled was particularly higher for small guidance compared to large guidance. When looking at the differences between informative and suggestive guidance, it was found that when receiving informative guidance, participants fulfilled a higher percentage of the decisional guidance when receiving large guidance. The reason for this could be that receiving a percentage as information instead of a suggested value highlights the size of the guidance. Participants could be triggered by such a big difference in their value and therefore are more likely to fulfill a higher percentage of the advice. However, this is not tested, and it would be interesting to further investigate this suggestion. When looking at the performance of decisional guidance with small advice versus big advice, this also shows that small advice leads to a lower $\mathrm{FVA}_{\mathrm{b,a}}$ (3.06% versus 22.26%). This can be seen as logical since the advice is bigger; therefore, your impact is also bigger. However, if the small advices of the decisional guidance add such little value, it must be further investigated whether it is worth it to use those. Receiving a lot of decisional guidance might also affect the willingness to fully accept certain decisional guidance. If that would result in planners not or, in a less extent taking over the guidance with big advice, it might not be worth using those small advices for decisional guidance.

Overall, decisional guidance can be a very effective way of improving the process of judgmentally adjusting forecasts. For items with medium to high volatility of demand, decisional guidance on the size can help planners in deciding on the right side of the adjustment. This can increase the added value of the adjustments since planners are better at deciding the direction of an adjustment than the size. For items with low volatility of demand, besides decisional guidance on the size, decisional guidance on the adjustment can be of added value by preventing planners from overreacting and nullifying the good performance of the statistical forecasts.

Chapter 8

Discussion

The previous chapter was devoted to presenting the conclusions based on the results of this thesis. In this chapter, the academic and practical implications are discussed. Furthermore, the limitations of this research are reviewed, and possibilities for future research are pointed out.

8.1 Academic implications

The academic goal of this thesis was to contribute to the existing literature on the topic of judgmentally adjusting forecasts and decisional guidance. A lot of research exists already on judgmental forecasting and also on the combination of statistical forecasts and judgmentally adjusting these forecasts. The existing literature on judgmentally adjusting forecasts has been changing over the years regarding the value of judgmental forecasting. Lawrence et al. (2006) concluded that in the early days, there was not yet the acceptance of the importance of judgment in forecasting. In later stages, there was a more general consensus on the importance of having judgment in forecasting (Lawrence et al., 2006). More in detail, it has been stated and proven the value that judgmentally adjusting forecasts can add to the forecast accuracy (Fildes et al., 2009; Franses & Legerstee, 2009; Arvan et al., 2019). This thesis contributes to the contradictory results that already exist in the literature on the added value of judgmentally adjusting forecasts. The results of the data used from EyeOn showed that, on average, planners did not improve but deteriorated the accuracy of the forecasts, which is not in line with the latest findings in the literature. However, this research also shows that in particular cases, the judgmental adjustments do add value to the forecast.

In the existing literature, there is limited to no research on the performance of judgmentally adjusting forecasts among different segmentation categories. This research shows where judgmental adjustments add value for forecast accuracy for the case of the customer of EyeOn. For items categorized as a Z-item, the adjustments did add value to the forecast accuracy. For X-items and Y-items, the judgmental adjustments led to a decrease in the forecast accuracy. These findings provide extra insights into the already existing literature on judgmentally adjusting forecasts.

In the field of judgmental forecasting, cognitive biases have been stated as one of the characteristics that could negatively affect the forecast performance of a planner when making adjustments (Sanders & Ritzman, 2004). There has been quite some research on the definition and working of different cognitive biases (Dietvorst et al., 2018; Fildes et al., 2007; Petropoulos et al., 2016). There has been less research on using measurements method to measure cognitive biases. Research of Pennings (2016) and (Eroglu & Croxton, 2010) did already focus on the individual behavior of planners in judgmental forecasting, thus on cognitive biases. This thesis showed that cognitive biases were also present among the customer of EyeOn and that individual differences were present as well. In the case of the overreaction bias, this can be directly linked to the performance of the forecast accuracy, with a negative impact since the definition of overreaction bias implies a deterioration of the forecast accuracy (Eroglu & Croxton, 2010). For the other two biases, there is not such a clear link to be found between the presence of the biases and performance, which is in line with existing literature (Fildes et al., 2009; Lawrence et al., 2006).

Regarding decisional guidance, there are studies on the different types and effectiveness in a more general way (Montazemi et al., 1996; Parikh et al., 2001) and to some extent in combination with forecasting (Fildes et al., 2006). The results of the conducted experiment contribute to the literature by combining decisional guidance with the judgmentally adjusting forecasts task. The results of the experiment in this thesis show, in line with the existing literature (Montazemi et al., 1996; Fildes et al., 2006; Parikh et al., 2001), a positive effect of the decisional guidance on the forecast performance. Research of Montazemi et al. (1996) concluded differences between the effectiveness of suggestive guidance and informative guidance depending on the complexity of tasks. In the experiment, the differences between the effectiveness of the forms of guidance have been limited. This is different from the existing literature (Montazemi et al., 1996), but this effect should be studied further to gain more knowledge in this aspect. It could be that the participants are an exception, and there are differences to be found in other studies. This thesis also contributed to the literature by testing the effects of suggestive guidance to not adjust the statistical forecast. The results on this aspect showed this type of decisional guidance had a positive effect on the forecast performance, especially for X-items. In literature (Silver, 1991; Fildes et al., 2006), restrictiveness has been stated as a dangerous option since it can frustrate the users. The suggestive guidance on the adjustment is not absolute restrictiveness, but it is a combination of guidance and restricting, and it would therefore be interesting to see the effects of this form on the frustration of users and trust in the system's guidance.

8.2 Practical implications

Besides the academic objective of this thesis, another objective was to provide more insights into the performance of the adjustments done at the customer of EyeOn and the possible effects that decisional guidance can have on forecast accuracy. The results of the data analysis show that the current situation and the performance of the adjustments to the statistical forecasts are far from optimal. The analysis shows that, currently, the adjustments, on average, do not add value to the statistical forecasts. On a more detailed level, the analysis shows that for X-items and, to a less extent, Y-items, this is especially the case. Adjustments made on Z-items are, on average, having a positive effect on forecast accuracy. In general, in most cases, the direction of the adjustment was right, but this does not always lead to an improvement in the forecast. The presence of the overreaction bias and the fact that the performance of the statistical forecast for Y-items and especially X-items are major reasons for these results.

The effects decisional guidance had in the experiment are not tested on data from the customer of EyeOn. However, in combination with the results of the data analysis, it does show the effect that decisional guidance can potentially have on EyeOn and its customers. For EyeOn, the main recommendation is to focus on how the decisional guidance should be ideally created. In the experiment, the decisional guidance was created based on assumptions that certain knowledge is known to the system. For EyeOn, it would be an opportunity to focus on creating opportunities to compute decisional guidance based on the earlier performance of the adjustments and historical information. Dynamic decisional guidance (Silver, 1991) can be a very interesting opportunity for EyeOn since there is enough data available, and it gives the opportunity to keep adding data to the used historical dataset. Furthermore, focusing on extending the dataset is another opportunity. Currently, the information about the adjustments is limited to basic information. Once this can be extended to, for example, reasons for adjustment, the data about the performance can be further specified and used as input for the decisional guidance. Currently, the option to fill in the reason for the adjustment is voluntary, and this makes it hard to use the data efficiently. If it were mandatory to select a reason out of, for example, the ten most common reasons to adjust, this would give a lot of additional interesting information. This can be done based on text analysis of the known reasons filled in by the planners or in cooperation with demand planners, managers, and other stakeholders. Other options to extend the data would be an analysis of which adjustment had the most impact on the forecast accuracy. This can be specified to adjustments of certain departments and to the different hierarchy levels.

Regarding the design of decisional guidance, for EyeOn, it would be beneficial to further focus on the design of the decision support system in cooperation with their customers. In the current situation for Company A, the adjustments can be made through Jedox. Due to the complexity of the usage of this system currently, some adjustments are made through different channels. It would be more convenient for analysis to make sure all adjustments are made through one platform. Therefore, it would be useful to focus on making the platform more user-friendly and increasing the convenience of use. Furthermore, the current alerts are not very remarkable, and it is unknown if planners really look at these alerts. Replacing the current alerts with decisional guidance and making the platform more user-friendly can improve the adjustment process and performance. When EyeOn will implement decisional guidance, it is recommended to use certain thresholds for when guidance is provided. In this master thesis during the experiment the decisional guidance was always given to the participant. However, this might lead to the fact that receiving the guidance is not that remarkable in the end. Therefore, it is good to carefully think about using the guidance when it can be most beneficial for the forecast accuracy.

8.3 Limitations

First of all, the data analysis is a case study which means that the results apply to this specific situation but not necessarily can be generalized to a much broader setting. The results of the data analysis come from one company which means the results are useful for their case, but this does not mean the same holds for other cases.

Due to the adjustments process conducted at Company A, the dataset does not contain all adjustments performed, and in some cases, the adjustments are not the final adjustment. This is due to the fact that an item can be adjusted multiple times in a period and because the platform in which the adjustments can be made is not very user-friendly, which leads to the fact that planners also make adjustments through different channels that are not logged in the same way. This might give an incomplete view of the situation, but it was the only possible way to use the adjustments from multiple hierarchy levels. This limitation is also in line with the recommendation to EyeOn to focus on improving the decision support system (Jedox) such that it will be easier in the future to gain insights on all adjustments made.

Regarding the experiment, there are also some limitations. First, it is an experiment which means it is different from a real-life situation. The situation that is created in the experiment uses two features as a trigger for the participants to make the decision if an adjustment is necessary. This is, of course, different than the situation planners have to deal with in real life. There are many more aspects (such as other departments, contractual agreements, and managerial influence) that planners need to take into account when deciding if an adjustment is necessary. This might also lead to the fact that planners in real life can react differently to decisional guidance. Therefore it is recommended to investigate the effects of decisional guidance further within a real adjustment process. This can be done with the implementation of decisional guidance in an existing FSS and testing it on real data or in an experiment with planners on historical data.

Furthermore, the experiment only contains 36 participants. It gives a good first insight, but more participants can be added to confirm that the results can be generalized.

In the experiment, decisional guidance has been calculated with the assumption that the system is able to detect the effect that features have on the actual demand. In real life, it would be difficult or even impossible to come to a situation in which this is exactly the case. The goal of this thesis was to investigate the effects of decisional guidance on the behavior of planners rather than creating the perfect decisional guidance. Therefore, it would be useful to focus on further research on the design and computational sides of decisional guidance and test these effects again to get a complete view of the effects.

8.4 Future research

In the previous sections, the academic and practical implications have been addressed, and the limitations of this research have been discussed. From these sections, the step to interesting opportunities can be assessed.

As stated above, it would be interesting for EyeOn, and academic contributions, to do research on the computational side of decisional guidance. Research should focus on what type of data can be used in order to create dynamic decisional guidance. For suggestive guidance, the focus could be on what data can be used best in order to create the most effective decisional guidance. For informative guidance, it would be more interesting to focus on the type of feedback that is most effective for improving the judgmental adjustments. There have already been multiple types of feedback discussed in the literature that could be used in a decision support system, such as outcome feedback, performance feedback, cognitive process feedback and task properties feedback (Fildes et al., 2006). From the results of this thesis, it would be an interesting next step to further investigate the different effects of these types of feedback on forecast performance.

The different effects of suggestive versus informative guidance would be an interesting topic to investigate further. In this research, the complexity has been represented by the segmentation of XYZ categories. Further research could be done on the complexity of the work environment. How complex tasks is can be determined by the number of stakeholders or other external aspects a planner needs to take into account. This work environment can, in general, be an interesting direction to incorporate in future research on decisional guidance in judgmental forecasting. In big companies, it is common for multiple departments to deliver a forecast as input for the final forecast. It would be interesting to research the effect of decisional guidance on these individual elements as well as on the task of combining these different inputs into one final forecast. Decisional guidance could, in that case, focus on the previous performance of each of the input forecasts.

Another interesting direction would be to use cognitive biases to investigate individual differences between the effects of decisional guidance. The results of this thesis did not find clear differences between the effects of decisional guidance in relation to cognitive biases. Therefore, it would be interesting to dive deeper into whether personal cognitive biases affect the working of decisional guidance. This could be done in combination with trying to make a link with forecast performance. With the current existing literature on cognitive biases, it is hard to make a direct link to forecast performance, so more extensive research on this combination would be beneficial for understanding this topic and obtaining practical insights. Furthermore, it can be interesting to investigate whether for different cognitive biases, different forms of decisional guidance work better.

References

Alvarado-Valencia, J., Barrero, L. H., Önkal, D., & Dennerlein, J. T. (2017). Expertise, credibility of system forecasts and integration methods in judgmental demand forecasting. *International Journal of Forecasting*, *33*(1), 298–313.

Arvan, M., Fahimnia, B., Reisi, M., & Siemsen, E. (2019). Integrating human judgement into quantitative forecasting methods: A review. *Omega*, *86*, 237–252.

Benson, P. G., & Önkal, D. (1992). The effects of feedback and training on the performance of probability forecasters. *International Journal of Forecasting*, *8*(4), 559–573.

Bolton, G. E., & Katok, E. (2008). Learning by doing in the newsvendor problem: A laboratory investigation of the role of experience and feedback. *Manufacturing & Service Operations Management*, *10*(3), 519–538.

Bowerman, B. L., O'Connell, R. T., & Koehler, A. B. (2005). *Forecasting, time series, and regression: an applied approach* (Vol. 4). South-Western Pub.

Cuppens, B. (2020). Exploring judgmental forecasting for a large company in the pc market.

Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to sku-level demand forecasts. *International Journal of Forecasting*, *29*(3), 510–522.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine learning research*, *7*, 1–30.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, *64*(3), 1155–1170.

Eroglu, C., & Croxton, K. L. (2010). Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting*, *26*(1), 116–133.

Eroglu, C., & Sanders, N. R. (2021). Effects of personality on the efficacy of judgmental adjustments of statistical forecasts. *Management Decision*.

EyeOn. (n.d.). *Eyeon: Planning services.* Retrieved from `https://eyeonplanning.com/offering/planning-services/`

Fildes, R., Goodwin, P., & Lawrence, M. (2006). The design features of forecasting support systems and their effectiveness. *Decision Support Systems*, *42*(1), 351–361.

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International journal of forecasting*, *25*(1), 3–23.

Fildes, R., Goodwin, P., et al. (2007). Good and bad judgement in forecasting: Lessons from four companies. *Foresight*, *8*(Fall), 5–10.

Franses, P. H., & Legerstee, R. (2009). Properties of expert adjustments on model-based sku-level forecasts. *International Journal of Forecasting*, *25*(1), 35–47.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, *19*(4), 25–42.

Goodwin, P. (2002). Integrating management judgment and statistical methods to improve short-term forecasts. *Omega*, *30*(2), 127–135.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, *22*(4), 679–688.

Keen, P. G. (1978). *Decisión support systems; an organizational perspective* (Tech. Rep.).

Kremer, M., Moritz, B., & Siemsen, E. (2011). Demand forecasting behavior: System neglect and change detection. *Management Science*, *57*(10), 1827–1843.

Lau, N., Hasija, S., & Bearden, J. N. (2014). Newsvendor pull-to-center reconsidered. *Decision Support Systems*, *58*, 68–73.

Lawrence, M., Goodwin, P., O'Connor, M., & Önkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of forecasting*, *22*(3), 493–518.

Li, Y., & Jiang, Q.-J. (2017). Demand forecasting and information platform in tourism. *Open Physics*, *15*(1), 247–252.

Montazemi, A. R., Wang, F., Nainar, S. K., & Bart, C. K. (1996). On the effectiveness of decisional guidance. *Decision Support Systems*, *18*(2), 181–198.

Montgomery, D. C., & Runger, G. C. (2010). *Applied statistics and probability for engineers*. John Wiley & Sons.

Moritz, B., Siemsen, E., & Kremer, M. (2014). Judgmental forecasting: Cognitive reflection and decision speed. *Production and Operations Management*, *23*(7), 1146–1160.

Nahmias, S., & Olsen, T. L. (2015). *Production and operations analysis*. Waveland Press.

Parikh, M., Fazlollahi, B., & Verma, S. (2001). The effectiveness of decisional guidance: an empirical evaluation. *Decision sciences*, *32*(2), 303–332.

Pennings, C. (2016). *Advancements in demand forecasting: Methods and behavior* (Tech. Rep.).

Petropoulos, F., Fildes, R., & Goodwin, P. (2016). Do 'big losses' in judgmental adjustments to statistical forecasts affect experts' behaviour? *European Journal of Operational Research*, *249*(3), 842–852.

Petropoulos, F., Goodwin, P., & Fildes, R. (2017). Using a rolling training approach to improve judgmental extrapolations elicited from forecasters with technical knowledge. *International Journal of Forecasting*, *33*(1), 314–324.

Ritzman, L. P., & King, B. E. (1993). The relative significance of forecast errors in multistage manufacturing. *Journal of Operations Management*, *11*(1), 51–65.

Sanders, N. R., & Ritzman, L. P. (2004). Integrating judgmental and quantitative forecasts: methodologies for pooling marketing and operations information. *International Journal of Operations & Production Management*.

Scholz-Reiter, B., Heger, J., Meinecke, C., & Bergmann, J. (2012). Integration of demand forecasts in abc-xyz analysis: practical investigation at an industrial company. *International Journal of Productivity and Performance Management*.

Schweitzer, M. E., & Cachon, G. P. (2000). Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science*, *46*(3), 404–420.

Silver, M. S. (1991). Decisional guidance for computer-based decision support. *MIS quarterly*, 105–122.

Singh, D. T. (1998). Incorporating cognitive aids into decision support systems: the case of the strategy execution process. *Decision support systems*, *24*(2), 145–163.

Worthen, B. (2003). Future results not guaranteed. *CIO-FRAMINGHAM MA-*, *16*(19), 46–53.

Appendix A

Statistical Methods

In this appendix the statistical methods used in this thesis to test the hypotheses are explained.

A.1 Student t-test

The student t-test can be used to determine the dependency between numerical variables. This method tests if the means of two variables are significantly different (Montgomery & Runger, 2010). The student t-test is based on a student t-distribution and its shape of the distribution relies on the degrees of freedom. The distribution is very similar to the normal distribution. It only varies in the fact that it does not have a set of standard deviations, but it applies a standard deviation of the applied data. With the sample standard deviation the significant difference can be tested.

A.2 Wilcoxon rank-sum test

The Wilcoxon rank-sum test (also known as the Mann-Whitney test) tests the null hypothesis that two samples are derived from the same distribution and therefore, have an equal mean. The two samples are independent and the sample sizes do not have to be the same for that reason. It ranks all the observations that have been made from smallest to largest and checks how many times an observation from one sample is ranked lower than an observation from the other sample. Scores from both samples are than put together and ranked, with the lowest score receiving rank one, the second lowest score rank two et cetera. Equal scores get the same rank assigned. The test statistic is the smallest of the sum of ranks for the two sets, called W. The p-value is derived from the probability that the larger rank is from the same distribution given the alternative of being unequal, higher or lower (Demšar, 2006).

A.3 Wilcoxon paired signed-rank test

The Wilcoxon signed-rank test calculates the differences between two paired samples. It uses two dependent samples and does not assume any distribution. It is based on the principal to rank the differences between two numerical samples. Each unique differences represents a rank. The sum of rank numbers is calculated and the p-value is based on the likeliness that the larger rank sum is unequal, larger or smaller than the small one, based on the alternative hypothesis.