Aalto University
School of Science
Department of Industrial Engineering and Management
Industrial Management

Teemu Belt

Master's thesis

When is forecast accuracy important in the retail industry? Effect of key product parameters

Master's Thesis
Espoo, 22.2.2017

Supervisor: Kari Tanskanen, Professor of Industrial Management

Instructor: Timo Ala-Risku, Doctor of Science in Technology

Abstract

Author: Teemu Belt		
Title: When is forecast accuracy important in the retail industry? Effect of key product parameters		
School: School of Science	Department: Industrial Engineering and Management	
Professorship: Industrial Management		Code: TU-22
Number of pages: 56	Date: 22.2.2017	Library location: TU

Accurate forecasting is important for retail companies that want to minimize the capital that is tied in stocks while simultaneously ensuring adequate product availability for their customers. Forecasting processes are typically automated; however, time-consuming and costly manual corrections to forecasts are often made. Therefore, retailers would benefit from clear criteria for selecting products for which forecast accuracy has the greatest impact. In addition, quantifying the business impact of forecast accuracy could help retailers that are considering investments in advanced forecasting solutions. Furthermore, researched and clearly documented evidence about the importance of forecast accuracy could aid communication both within retail companies and with other players in the supply chain. For these reasons, the goal of this thesis is to identify the situations in which forecast accuracy is important and those in which it is not.

This thesis utilizes real sales figures from a major European retailer and mathematical simulations to clarify when forecast accuracy is important. Approximately 7 million product locations were included in the thesis, and the data covered sales history for 24 months. With access to real product-specific sales data, the researcher was able to simulate the future using real sales figures from the past. In other words, it was possible to go a few years back in time and test the impact a certain forecast would have had. Using real sales and product data, it was possible to vary the key forecast and product location parameters and then to simulate the business outcome.

The results of this study suggest clear principles for prioritizing product locations and quantify how much additional stock is needed to compensate for forecast errors. The results also reveal the impact of different product location parameters on the business importance of forecast accuracy. The *sales volume* of a product location proved to be the most important parameter, although *relative batch size* also had some importance. *Average time to delivery* plays a minor role in some cases, while *relative sales standard deviation* (STD) had only an insignificant effect.

Based on the results of this thesis, retailers should consider focusing their efforts on improving the forecast accuracy of high-selling products and, to some extent, on products with small batch sizes. Average time to delivery deserves closer scrutiny only in cases with systematic forecast errors. Relative sales STD proved so marginal that retail managers could consider ignoring it.

Supervisor: Kari Tanskanen, Prof. Instructor: Timo Ala-Risku, D.Sc (Tech.)
Keywords: Forecasting, retail business, statistical methods

Tiivistelmä

<table>
<tr><td colspan="3">Tekijä: Teemu Belt</td></tr>
<tr><td colspan="3">Työn nimi: Milloin ennustetarkkuus on tärkeää vähittäiskaupassa? Tuotteen avainparametrien vaikutus</td></tr>
<tr><td>Koulu: Perustieteiden korkeakoulu</td><td colspan="2">Laitos: Tuotantotalous</td></tr>
<tr><td colspan="2">Professuuri: Teollisuustalous</td><td>Koodi: TU-22</td></tr>
<tr><td>Sivuja: 56</td><td>Päiväys: 22.2.2017</td><td>Työn sijainti: TU</td></tr>
<tr><td colspan="3">Tarkat ennusteet ovat tärkeitä vähittäiskaupan yrityksille, jotka haluavat minimoida varastoihin sitoutuneen pääoman, mutta samalla pitää riittävästi tuotteita saatavilla asiakkailleen. Ennustaminen on tyypillisesti pääosin automatisoitua, mutta ennusteisiin tehdään usein käsin kalliita ja aikaa vieviä korjauksia. Vähittäiskauppa hyötyisi selkeistä kriteereistä, joilla valita ne tuotteet, joille ennustetarkkuus on kaikkein tärkeintä. Ennustarkkuuden vaikutusten kvantifioiminen voisi olla höydyllistä myös silloin, kun vähittäiskauppa harkitsee investointeja edistyneempiin ennustemenetelmiin. Lisäksi selkeästi esitetty tutkimustieto ennustetarkkuuden merkityksestä edistäisi kommunikointia vähittäiskaupan toimitusketjussa. Näistä syistä tämän diplomityön tavoitteena on tunnistaa, milloin ennustetarkkuudella on merkitystä ja milloin ei.

Diplomityön tavoitteen saavuttamiseksi hyödynnettiin yhden merkittävän eurooppalaisen vähittäiskauppaketjun myyntitietoja, joiden pohjalta tehtiin simulointeja. Data sisälsi noin 7 miljoonaa tuotelokaatiota, joiden myyntihistoriaa tarkasteltiin 24:n kuukauden ajalta. Toteutuneen myynnin avulla oli mahdollista mennä ajassa taaksepäin ja tutkia, millaisia vaikutuksia erilaisilla ennusteilla olisi ollut. Käyttäen todellista myynti- ja tuotedataa oli mahdollista simuloida ennuste- ja tuoteparametrien vaikutusta liiketoimintaan.

Tämän tutkimuksen tulokset tarjoavat selkeän tavan priorisoida tuotelokaatioita ja kvantifioida sitä, kuinka paljon lisää varastoa tarvitaan ennustevirheen kompensointiin. Tulokset kuvaavat myös, miten eri tuotelokaatioparametrit vaikuttavat ennustevirheen liiketoiminnalliseen merkitykseen. Tärkeimmäksi parametriksi osoittautui myyntivolyymi ja verrattain tärkeä oli myös suhteellinen eräkoko. Keskimääräisellä ajalla seuraavaan toimitukseen oli jossain tilanteissa pieni vaikutus, mutta myynnin hajonnalla ei ollut merkittävää vaikutusta.

Tulosten perusteella vähittäiskaupan kannattaisi keskittää ennustetarkkuuden parantamiseen tähtäävät panostukset tuotteisiin, joita myydään paljon, sekä jossain määriin tuotteisiin, joilla on pieni eräkoko. Keskimääräinen aika seuraavaan toimitukseen on tarkkailemisen arvoinen vain tilanteissa, joissa ennusteessa on systemaattista virhettä. Myynnin hajonnan vaikutus on niin pieni, että vähittäiskauppiaat voivat harkita sen jättämistä pois tarkastelusta.</td></tr>
<tr><td colspan="3">Valvoja: Prof. Kari Tanskanen
Ohjaaja: TkT Timo Ala-Risku</td></tr>
<tr><td colspan="3">Avainsanat: ennustaminen, vähittäiskauppa, tilastolliset menetelmät</td></tr>
</table>

Key terms

Product location	A specific product in a specific location (store)
Forecast error	Difference between actual sales and forecasted sales
Forecast bias	Tendency for forecasts to be too high (positive bias) or too low (negative bias)
Random forecast deviation	Random deviation of forecast from actual sales
Availability	Percentage of days that end with a stock of more than zero
Inventory days of supply[1] Relative inventory[2]	How many days of average sales it takes to sell all items in inventory; average stock level divided by average daily sales
Supply of order batch size[1] Relative batch size[2]	How many days with average sales it takes to sell one batch; batch size divided by average daily sales
Batches sold per day	Average number of batches sold per day
Lead time[1] Delivery time[2]	Time between order and delivery
Order interval	Number of days between two possible order dates
Average time to delivery	Expected time until the next possible delivery; $(Lead\ time + \frac{order\ interval}{2})$
Average daily sales	Average daily sales
Relative sales standard deviation	Standard deviation of sales divided by average daily sales

[1] Established RELEX term

[2] Alternative term

Contents

1 Introduction

1.1 Motivation for the study

Forecasting is important for retail companies, which, like all firms, want to minimize the capital tied in their stocks (Bowersox, Closs, & Cooper, 2002; Levy & Weitz, 2011). Additionally, there are retail-specific reasons for demanding high forecast accuracy, including relatively short product life-cycles and high cost of space in popular store locations (Bowersox et al., 2002; Stadtler & Kilger, 2004).

The forecasting process is mostly automated; however, manual corrections to the forecasts and improvements to the forecasting models are often necessary (Fildes & Goodwin, 2009). Manual corrections are time-consuming and costly, and therefore it is crucial to know when improved forecast accuracy is worth the effort. Consequently, retailers would benefit from clear criteria for selecting products for which forecast accuracy has the greatest impact.

Quantifying the business impact of forecast accuracy could also help retailers that are considering investments in advanced forecasting solutions. Furthermore, knowing when forecast accuracy is not important might help avoid unnecessary efforts to improve forecast accuracy.

In addition, having clear, researched evidence about the importance of forecast accuracy could aid communication both within retail companies and with other players in the supply chain. Clear and illustrative communication is important as forecast-related topics are complicated, the amount of data is hard for humans to grasp, and there are many people involved. This is often the case in the retail industry, in which large companies may have millions of product locations and dozens of people working on tasks related to forecasts.

1.2 Research goal

For the reasons discussed above, the goal of this thesis is to identify the situations in which forecast accuracy is important and those in which it is not. To answer that, the interdependences of forecast accuracy, context, and supply chain performance need to be analyzed. Figure 1 illustrates these conceptual interdependences.

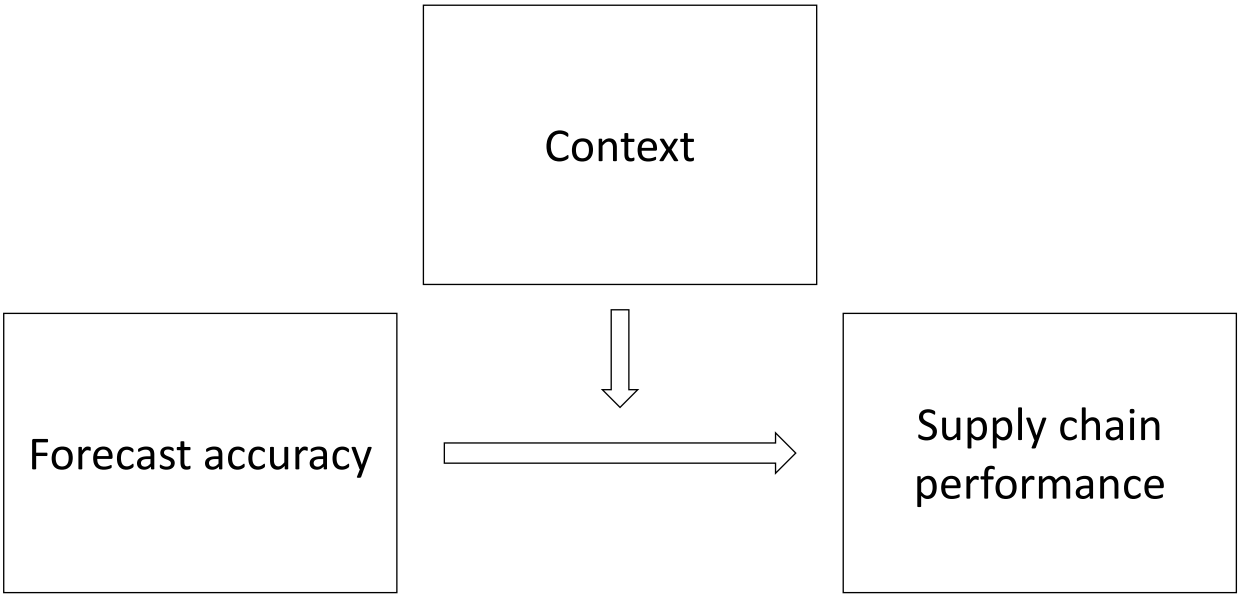

Figure 1. Conceptual interdependences of forecast accuracy, context, and supply chain performance

The research goal will be discussed in detail in Section 1.4 (Research questions) after the gaps in the scientific literature have been discussed.

1.3 Gaps in previous research

In the scientific literature, there is a lot of discussion about the effects of forecast accuracy. However, some aspects of this topic have received little attention. First, the scientific literature does not adequately cover the importance of forecast accuracy in the retail sector. Most studies focus on manufacturing (Biggs & Campion, 1982; Fildes & Kingsman, 2011; Ritzman & King, 1993) or warehouse environments (Sanders & Graman, 2009), whereas this thesis focuses on the retail environment. Second, scientific studies often use randomly generated demand data (Biggs & Campion, 1982; Ritzman & King, 1993; Xie, Lee, & Zhao, 2004). It seems that scientific literature seldom uses real business data, such as sales history. Third, the factors affecting the importance of forecast accuracy have not been adequately covered. Specifically, the scientific literature seems to lack systematic and quantified descriptions of how different parameters affect the importance of forecast accuracy. Section 1.4 will describe how this thesis aims to fill these gaps in previous research.

1.4 Research questions

As illustrated earlier in Figure 1, understanding the interdependences of forecast accuracy, context, and supply chain performance is important in this study. Figure 2 illustrates these interdependences when applied to the context of the study. Forecast accuracy is represented by measurable metrics: *forecast bias* and *random forecast deviation*. Context is represented by product location parameters: *batch size*, *lead time*, *order interval*, *average daily sales* and *relative sales standard deviation (STD)*. Supply chain performance is represented by key performance indicators (KPIs): *availability* and *inventory days of supply*. All the aspects highlighted in italics above can be expressed numerically, which supports the purpose of this study.

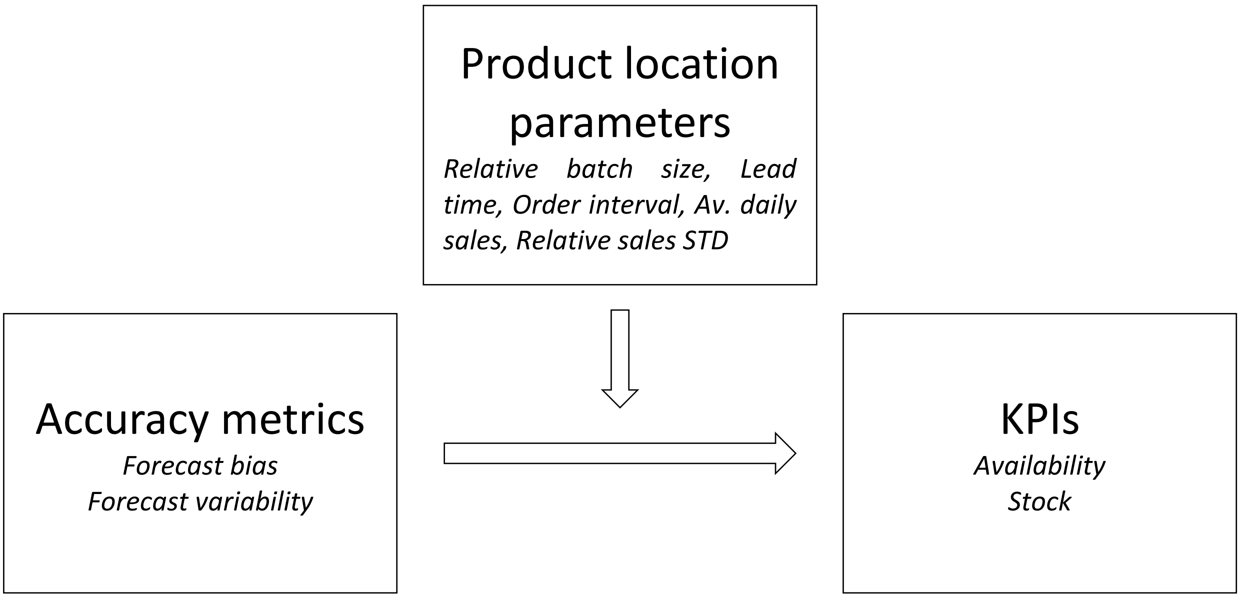

Figure 2. The interdependences of forecast accuracy, context, and supply chain performance applied to this thesis

The interdependences of forecast accuracy, context, and supply chain performance can be expressed using the following three research questions:

RQ1. What is the relationship between availability and stock?

In order to measure supply chain performance, relevant KPIs must be defined. In the retail context, these include 1) product availability at point of sales and 2) required stock. As described by Johnston, Taylor, & Oliveria (1988), these two KPIs are interdependent, and in order to accurately estimate the overall supply chain performance, both of them need to be taken into account.

RQ2. How does forecast error affect supply chain performance?

Intuitively, it would make sense for high forecast error to worsen supply chain performance, but this might not always be the case. Additionally, the sensitivity to forecast error should be quantified into a variable in order to be used in further analyses. This master's thesis follows Hausman (2004), who suggests that overall supply chain performance could be assessed by analyzing product availability, required stock, and the interdependence between these two factors.

RQ3. How do different product location parameters affect sensitivity to forecast error?

By studying how sensitivity to forecast error is affected by product location parameters, such as lead time and batch size, it is possible to identify when forecast accuracy is important.

1.5 Research environment

The research background of this master's thesis includes real sales figures from a major European retailer and the supply chain management tools of RELEX Solutions. Approximately 7 million product locations were included, and the data covers sales history for two years. In order to keep the scope of this master's thesis reasonable, only data from a single case were included.

All the main product groups of the chosen retailer were included in the data to be analyzed. However, only products with relatively long shelf lives were included, and quickly spoiling food products were excluded as modeling the spoilage of products differs significantly from modeling other aspects of supply chain optimization.

Supply chain management software was used to run mathematical simulations to clarify the situations in which forecast accuracy is important. The software includes features for demand forecasting, automatic replenishment, and inventory management. This master's thesis was conducted at and for RELEX Solutions, which provided the retailer sales data.

1.6 Methodology of the study

1.6.1 A quantitative approach

The purpose of this thesis is to clarify the situations in which forecast accuracy is important and define clear criteria for selecting products for which forecast accuracy will have the greatest business impact. In order to prioritize products for which forecasts need to be manually corrected, their priority should be quantified. Similarly, in order to properly consider investments in new forecasting solutions, their benefit should be quantified. Communication is also easier when backed by clear numerical results. These arguments support the choice to use a quantitative approach for this thesis. In addition, there was a possibility to use real sales figures in a quantitative form.

1.6.2 Real sales data

The word "forecast" refers to the future; however, this study takes advantage of the possibility of simulating the future using real sales figures from the past. Using actual sales history, it was possible to study how different forecasts affect supply chain performance. In other words, it was possible to go a few years back in time and test the impact that a certain forecast would have had. In this thesis, the two key forecast parameters, bias and variability, are varied in order to identify how they affect business outcomes.

1.6.3 Product location parameters

The above-mentioned forecast parameters, bias and variability, can be influenced by the forecaster (e.g., RELEX). Conversely, product location parameters, such as batch size, lead time, and time between orders, are typically determined by retailers or their supply chains, not by the forecaster. In this master's thesis, the forecast parameters that can actually be influenced by the forecaster are varied. In doing so, this master's thesis aims to study the impact of real-life variations in product parameters using the available business data.

2 Literature review

2.1 Retail business

2.1.1 Key functions of retail

Retail can be defined as selling products to end-users (i.e., consumers) (Levy & Weitz, 2011). However, it should be noted that retailers can also sell to businesses (Ayers & Odegaard, 2007). Retailers' role in a supply chain is to link manufacturers to consumers (Levy & Weitz, 2011).

Retailers can create value for their customers in several ways. Levy & Weitz (2011) divide these value-creating activities into four categories. First, retailers provide a variety of products in one store so that consumers can satisfy their purchasing needs without visiting multiple stores. According to Ayers & Odegaard (2007), this is the most important function of retailers. Second, retailers buy products in large quantities at low prices from manufacturers, or wholesalers, and sell them to consumers (Levy & Weitz, 2011). This arrangement serves the needs of both the consumers, who want to buy products in small volumes, and manufacturers, who specialize in producing large volumes of products at low costs. Third, retailers maintain an easily accessible inventory for consumers and avoid storing large quantities of products themselves. Fourth, retailers provide services in addition to products, such as displaying products for customers to see and test, providing options to buy products on credit, offering warranties, and gift wrapping.

2.1.2 Availability and inventory in retail business

Creating value for customers usually requires having products in stock. The ability to have inventory when it is needed by a customer is called *availability* (Bowersox et al., 2002). In the retail context, this refers to the availability of products to end users at the point of sale. Agrawal & Smith (2015) argue that not having a product on a shelf often causes customers to leave the store.

It is, obviously, possible to increase *availability* by increasing *inventory* (Fildes & Kingsman, 2005). However, this is not always the best option, as maintaining a limited stock is important for a retailer's financial success (Levy & Weitz, 2011). According to Bowersox, Closs, & Cooper (2002), good store locations are costly, and therefore it is logical to minimize the area required for the store and the number of products in stock.

Value-creating activities are related to inventory in various ways (Levy & Weitz, 2011). In order to offer products to customers at any time, products must be available in the store's inventory. Offering a large variety of products increases the total number of products in stock. Even if a small number of each product is kept in stock, the total stock can be significant when there are tens of thousands of different products. Additionally, buying in large batches causes temporary peaks in stock levels. Furthermore, providing services, such as the option to see and test products, may require more products to be in stock.

2.2 Key performance indicators

There are a few key aspects of supply chain operations that determine the performance of the chain (Hausman, 2004), including product availability, inventory level, and costs related to personnel, logistics, and procurement. There are various metrics for measuring these aspects. These metrics are not always independent from each other; there are often tradeoffs between two or more metrics (Fildes & Kingsman, 2005). For example, it is possible to increase *availability* by increasing *inventory level*. These tradeoffs can be modeled mathematically as response curves (Johnson, Tellis, & Ip, 2013). The relationship between *stock level* and *availability* is especially important (Bowersox et al., 2002). Due to the high cost of maintaining inventory, it is important for retailers to minimize their stock levels. However, they also need to keep their customers happy by providing a sufficient service level.

There are several ways to measure availability, but two commonly used ones are *cycle service level* and *fill rate* (Syntetos, Nikolopoulos, & Boylan, 2010). *Cycle service level* can be defined as the percentage of replenishment cycles in which there are no stock-outs, that is, all demand can be satisfied directly from the stock. *Fill rate* can be defined as the percentage of total demand that can be delivered from the stock.

Calculating cycle service level or fill rate can be challenging in practice. In a case in which a product's stock is zero, it may be impossible for the retailer to know whether there is demand for the products. For this reason, RELEX measures availability as the percentage of days that end with more than zero stock. This can be seen as an approximation of cycle service level, and it was the way in which availability was measured in this thesis.

2.3 Measuring forecast accuracy

According to Krajewski, Ritzman, & Mallhotra (2013), forecasts are predictions of the future that are made for planning purposes. There are different types of forecasts, such as demand, sales, capacity, and backflow forecasts (Lu, 2014). In retail, demand forecasts are used to calculate optimal orders (Ayers & Odegaard, 2007). This is done to minimize inventory and ordering costs while avoiding stock-outs.

Forecasts can be based on mathematical models, managerial judgment, or a combination of both (Krajewski et al., 2013). Forecasters try to select the most accurate method for each situation. However, even with the best methods, forecast error can still occur (Biggs & Campion, 1982).

Forecast error is usually defined as the difference between the actual and forecasted demand (Krajewski et al., 2013). This can be expressed as a formula (Chase, 2013):

$$e_t = a_t - f_t \tag{1}$$

where e_t *= forecast error for period t*
a_t *= demand for period t*
f_t *= forecasted demand for period t*

2.3.1 Forecast error metrics

Formula (1) only defines error for a single period. To summarize forecast accuracy over several periods, different metrics must be used (R. J. Hyndman & Koehler, 2006). One simple approach is to use the mean absolute deviation (MAD) (T. Lee, Cooper, & Adam, 1993):

$$MAD = \frac{1}{n}\sum_{t=1}^{n}|a_t - f_t| \tag{2}$$

However, there is no one metric that is suitable for every situation; choosing an applicable forecast error metric should depend on the nature of the demand data. When choosing a metric, it is crucial to consider whether it is scale-independent (R. Hyndman, 2014). For instance, a change expressed in absolute units is scale-dependent, while a relative change is scale-independent. A scale-independent metric yields the same results regardless of scale (for example, using thousands of units instead of units sold) (Leitch & Tanner, 1991). Scale-independent metrics should be used when comparing different data series, such as the demand forecasts of different product locations, as is the case in this thesis (Krajewski et al., 2013). For this reason, scale-independent metrics were used here.

One scale-independent metric is *mean absolute percentage error* (MAPE), defined as follows (Ott, Mensendiek, & Gmeinwieser, 2013):

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{a_t - f_t}{a_t}\right| \tag{3}$$

There are some problems with MAPE that make it unsuitable for the purposes of this thesis (Ott et al., 2013). For instance, if a product does not sell during a given period, a division by zero occurs, creating undefined or infinite values. Similarly, low sales can lead to very high values. When comparing individual product locations, it is common to observe zero sales in a given period, and therefore a different metric should be chosen.

Weighted absolute percentage error (WAPE) largely solves the above-mentioned problem of undefined values and provides consistent measurements across different product locations (Hoover, 2009). Instead of dividing the error by the sales for each period and summing them up, it sums all the errors together and divides it by the sum of sales, thus avoiding the division by zero. The calculation logic is shown in the following formula:

$$WAPE = \frac{\sum_{t=1}^{n}|a_t - f_t|}{\sum_{t=1}^{n} a_t} \tag{4}$$

WAPE is undefined only when the total sales are zero, a very rare case that is unimportant. Additionally, WAPE is advantageous because it is relatively easy to understand (Hoover, 2009); the scientific literature also discusses other metrics that are more complicated to interpret (Cleophas, Frank, & Kliewer, 2009; R. Hyndman, 2014).

2.3.2 Forecast bias and random deviation metrics

WAPE measures total forecast error. However, as pointed out by Zhao, Xie, & Wei (2002) and Sanders & Graman (2009), forecast bias and random deviation may have different effects on supply chain performance. For this reason, it would be beneficial to measure these factors separately. Forecast bias occurs when forecasts are systematically too high or too low, whereas random deviation occurs when the forecasted value differs from the mean forecast (Biggs & Campion, 1982).

Forecast bias can be measured using the cumulative sum of forecast errors (CFE), sometimes defined as follows (Krajewski et al., 2013):

$$CFE = \sum_{t=1}^{n} (a_t - f_t) \tag{5}$$

This metric has the problem of being scale-dependent, but it can be made scale-independent by dividing it by total sales:

$$Relative\ Cumulative\ sum\ of\ Forecast\ Errors\ (RCFE) = \frac{\sum_{t=1}^{n}(a_t - f_t)}{\sum_{t=1}^{n} a_t} \tag{6}$$

According to Ritzman & King (1993), random deviation in forecasts can be measured using the standard deviation of the forecast error, as presented in Formula (7):

$$Forecast\ standard\ deviation\ (FSD) = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(a_t - f_t)^2} \tag{7}$$

This metric is also scale-dependent, but it can easily be made scale-independent by dividing it by average sales:

$$Relative\ forecast\ standard\ deviation\ (RFSD) = \frac{\sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(a_t - f_t)^2}}{\frac{1}{n}\sum_{t=1}^{n} a_t} \tag{8}$$

2.3.3 Summary of forecast error metrics

The forecast error metrics used in this thesis are summarized in Table 1. WAPE was used as in previous literature, whereas RCFE and RFSD, highlighted with a grey background, had to be modified to fit the purposes of this thesis.

Table 1. Summary of forecast error metrics

Forecast error type	Metric name	Alternative names	Formula	Source
Total error	WAPE	MAD/mean ratio, MAPE	$\frac{\sum_{t=1}^{n} \lvert a_t - f_t \rvert}{\sum_{t=1}^{n} a_t}$	(R. Hyndman, 2014)
Bias	RCFE		$\frac{\sum_{t=1}^{n} (a_t - f_t)}{\sum_{t=1}^{n} a_t}$	Modified from (Krajewski et al., 2013)
Random deviation	RFSD		$\frac{\sqrt{\frac{1}{n-1}\sum_{t=1}^{n} (a_t - f_t)^2}}{\frac{1}{n}\sum_{t=1}^{n} a_t}$	Modified from (Ritzman & King, 1993)

2.3.4 Importance of measuring forecast accuracy

Measuring forecast accuracy is important for several reasons (Kerkkänen, Korpela, & Huiskonen, 2009). First, forecast error measurements can be used to correct systematic forecast errors. For example, a forecast that is systematically too low can be adjusted to be higher. Second, forecast error metrics can be used to assess the impact of manual corrections made to a forecast (Chase, 2013). Third, different statistical models use forecast error measures to determine which forecast fits the historical demand or best predicts future sales (Chase, 2013). Fourth, even in cases in which forecasts cannot be improved, forecast error measurements can be used to plan how to cope with errors (e.g., by setting appropriate safety stocks) (Kerkkänen et al., 2009).

The forecast accuracy metrics discussed above do not measure the economic impact of forecast errors (Leitch & Tanner, 1991; Ott et al., 2013). The connection between forecast accuracy and economic impact is discussed in the following sections.

2.4 Effects of forecast accuracy

2.4.1 Is accuracy worth it?

Companies want to have the most accurate forecasts possible. However, more accurate forecasts tend to cost more both during implementation and use (Biggs & Campion, 1982). Therefore, obtaining the best forecast accuracy might not always be feasible (Fildes & Kingsman, 2011). In some cases, it might be more effective to improve supply chain performance by means other than improving forecast accuracy (Biggs & Campion, 1982). For example, it might be better to reduce lot sizes or increase flexibility. For these reasons, some kind of cost–benefit analysis is needed.

An ideal approach for choosing a forecasting system is to minimize the total cost. This includes both the cost of creating the forecast and the cost of forecast errors. Unfortunately, this is not always possible in reality due to the difficulty of estimating the cost of forecast errors. However, even roughly estimating these costs might help prioritize improvement efforts (Kerkkänen et al., 2009).

2.4.2 Effects of forecast accuracy

According to Wemmerlöv (1989), an inventory system operating with imperfect forecasts is fundamentally different from one with perfect forecasts. The most obvious difference is the former system's inability to satisfy all demand. Forecast errors also have other impacts (Kerkkänen et al., 2009), most notably on stock, availability, and total costs (Biggs & Campion, 1982).

The literature discusses different efforts to quantify the effect of forecast errors on total costs or other performance metrics. The reported results have been mixed. Fildes & Kingsman (2011) suggest exponentially increasing costs to maintain the same service level when forecast error increases. However, Sanders & Graman (2009) claim that costs increase linearly with an increase in forecast error.

As suggested by Barman, Tersine, & Burch (1990); Kerkkänen, Korpela, & Huiskonen (2009); and Sanders & Graman (2009), forecast *bias* and *random deviation* might have different effects on supply chain performance.

According to the literature, there may be a need to model positive and negative biases separately as they might behave differently (Barman et al., 1990). Positive forecast bias can cause excess stocks, excess manpower, and a need for discount prices (Biggs & Campion, 1982; Ott et al., 2013), while negative bias can lead to lost sales and a shortage of manpower.

Interestingly, deliberate biasing of forecasts can sometimes have a positive effect on system performance (T. S. Lee & Shih, 1989). However, as pointed out by Sanders & Graman (2009), it is not a practical option as it has the obvious risk of biasing the forecast too much and thus damaging performance. In addition, Sanders & Graman did not find the improvements to be statistically significant in many cases.

2.5 When is forecast accuracy important?

2.5.1 Motivation to study situations in which forecast accuracy is important

Forecast accuracy has an impact on supply chain performance, but the impact is not the same in all situations (Ritzman & King, 1993). As pointed out by Hoover (2009), some products are more important than others when it comes to forecast accuracy.

Hoover (2009) suggests assigning different weights to different products to reflect their differing economic impacts. This kind of prioritization, in combination with forecast accuracy metrics, might allow a forecaster to identify the most important products. Ott et al. (2013) reports that a model implemented at a large German retailer demonstrated that a scoring model is easy to implement, flexible, and can provide important insights for improving demand planning processes.

Prioritizing products based on the impact forecast errors have on them is especially interesting from the viewpoint of potential manual adjustments. The number of products is often so high that demand forecasting has to be automated using statistical methods. However, these automated forecasts can be improved by manual adjustments made by experts (Syntetos et al., 2010). A human expert can correct faults in the applied statistical model or consider circumstances that were not included in the model (Fildes & Goodwin, 2009). This process of automatically creating an initial forecast and then manually adjusting it is common in companies. These adjustments can lead to substantial savings regarding inventories (Syntetos et al., 2010). However, they also create substantial costs and slow down the creation of forecasts (Fildes & Goodwin, 2009). Therefore, there is a need to analyze the importance of forecast accuracy at the level of individual products.

The importance of forecast accuracy for a product is affected by factors such as sales volume (Kerkkänen et al., 2009) and lot size (Ritzman & King, 1993). Analyzing these factors might help companies calculate products' priorities and then focus their forecasting efforts on the products that matter the most.

2.5.2 Possible factors affecting the importance of forecast accuracy

The scientific literature identifies several product location parameters that affect the importance of forecast accuracy.

According to Ritzman & King (1993), batch size could act as a buffer, similar to safety stock, and thus affect the importance of forecast accuracy. Also, Fildes & Kingsman (2011) suggest that reducing lot sizes would lead to increased sensitivity to forecast errors.

Ott et al. (2013) points out that lead time might also have an effect as it determines how fast a shortage of a product can be fixed. For example, if the forecast for a product with a long lead time was too low, stock-outs could occur, but if the lead time is very short, more products could be purchased quickly and the stock-out would be brief.

Hoover (2009) argues that sales volume could be a significant factor affecting the importance of forecast accuracy (see also Ott et al. 2013). Further, Kerkkänen et al. (2009) argue that it is not feasible to spend much effort on improving forecasts for low-volume products. If the volume of a product is low, lot size is considered more important than forecast accuracy. Stock-out of a high-selling product results in many customers experiencing bad service.

Uneven demand might also affect the importance of forecast accuracy (Ritzman & King, 1993). Fildes & Goodwin (2009) suggest measuring demand unevenness using a *coefficient of variation* (i.e., the standard deviation of sales divided by average sales).

3 Research method

3.1 Choosing an appropriate method

There were two basic quantitative options to choose from: simulations and analytical solutions. As pointed out by Fildes & Kingsman (2011), analytical solutions tend to require simplifying the problem in order to make the mathematics manageable. According to Lu (2014), these simplifications might result in impractical solutions, even though some insights into the supply chain might be obtained.

In simulations, demand can either be randomly generated or based on real sales. However, randomly generated demand can cause some problems regarding the reliability of the results. For example, in the scientific literature, it is common to assume normally distributed demand, but this can lead to negative demand, a situation that should not occur in reality (Strijbosch & Moors, 2005).

It is often not possible to know the real demand for a product (Cleophas et al., 2009). A retailer knows the real sales, but, for example, in the case of a stock-out, sales differ from demand. However, sales data can serve as a sufficient approximation of demand (Syntetos et al., 2010).

This thesis used the simulation approach and real sales data to approximate demand. Simulation results were analyzed using the statistical methods described in the following subsection.

3.2 Statistical methods

Regression is a statistical method that describes how a variable depends on another variable or variables (Freedman, Pisani, & Purves, 2007). For example, *stock level* might be used to explain availability. In this example, stock level would be an independent variable, and availability would be the dependent variable. If there is more than one independent variable, the method is called "multiple regression analysis" (Mendenhall & Sincich, 2011).

Ordinary least squares (OLS) is a commonly used method of regression (Sheather, 2009). It assumes a linear relationship between the dependent and independent variables, as described in Formula (9):

$$y = a + kx + e \tag{9}$$

where $y =$ *dependent variable*
$a =$ *constant*
$k =$ *slope*
$x =$ *independent variable*
$e =$ *error term*

A regression model seldom explains the dependent variable exactly, and therefore an error term, e, is included in the formula above. OLS aims to minimize the sum of the squared errors.

In multilevel regression modeling, regression coefficients are estimated using another regression model. Multilevel models are well suited for situations in which the data are organized in groups. A common solution is to use a two-level regression model (Gelman & Hill, 2007). In the first level, regression coefficients are estimated for each group. Then, in the second level, the estimated coefficients are explained using a group-level variable.

3.3 Research process

This thesis investigated the importance of forecast accuracy in retail companies by simulating forecasts using real customer data. Numerous forecasts with deliberate variations were created and tested to determine how different forecast errors affect business outcomes. Figure 3 illustrates the research process utilized in this master's thesis. Each of the phases of the research process will be discussed in the following subsections. The conclusions phase will be discussed in Section 6 (Discussion and conclusions).

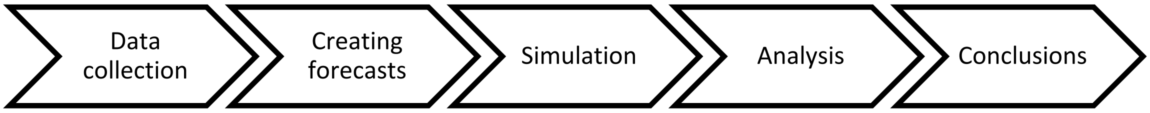

Figure 3. Research process

The data used in this thesis consists of real sales history and product information from a European retail company, an important customer of RELEX. Different forecasts were created by introducing intentional errors into the sales data. During the simulation phase, orders were calculated in the same way as they would be for real retail replenishment, using the initial stock level and different deliberately modified forecasts. Real sales history was used as demand to calculate changes in inventory. During the analysis phase, the impacts of different forecast errors and product location parameters on average stock level and availability were analyzed. Finally, conclusions were made regarding when forecast accuracy is important.

3.4 Data collection

The data used in this thesis consists of real sales history and product information from the case company. This customer company has approximately 7 million product locations, and the data included sales from 1 May 2014 to 30 April 2016. Real sales data were used to estimate actual demand and generate different forecasts to be used in simulations.

All the main product groups of the chosen customer were included in the data to be analyzed. These product groups included bakery products, baking supplies, beverages, cleaning supplies, cosmetics and hygiene products, dairy, frozen food, fruits and vegetables, meat and fish, ready-made meals, sweets, and snacks. Only products with relatively long shelf lives were included.

3.5 Creating forecasts for simulations

With the real demand known, differing "forecasts" were created for situations that occurred a few years earlier by introducing deliberate deviations from the actual demand. Using this approach, it was possible to study the effects of perfect forecasts on supply chain performance implications by using the realized demand as the forecast. Imperfect forecasts could be studied by introducing intentional errors into the perfect forecast. This approach had some advantages. First, it allowed the creation of perfect forecasts, something that is impossible in real business. Second, forecast biases and random deviations could be created separately from each other. Third, forecast errors could be created precisely in the desired quantities.

This master's thesis analyzed forecasts with numerous biases, random deviations, and combinations of both. In total, 11 levels of forecast bias and 7 levels of random forecast deviation were used. This resulted in a total of 77 different forecasts, as all combinations of biases and random deviations were tested.

3.6 Simulations

This thesis was based on simulations run using RELEX software. Using initial stock, demand, and forecast values as inputs, the software was used to calculate the key operational parameters, such as orders, deliveries, and stock levels. The resulting availability levels, stock levels, and other results were exported for further analysis, which was carried out using the statistical software R.

Numerous simulations were run using different levels of safety stock and forecast error. In total, 77 forecast error levels and 21 safety stock levels were analyzed, resulting in 1,617 simulations. The simulations took over 40 hours of server time to run, and related analyses using R took over 10 hours. The number of simulations proved adequate for the purposes of this study, and the required server time did not significantly harm ordinary business operations.

All product parameters that are determined by the case customer's business environment and cannot be changed by RELEX, including lead time, batch size, and order interval, were kept constant. Also, the demand data remained the same in all simulations. Even though these product location parameters were kept constant in all simulations, the business data used as input included variations in these parameter values. The impact of these variations was analyzed as a separate step.

The statistical mathematics used for analysis makes no distinction between parameters that RELEX can control and those that are determined by external factors. However, the interpretation of this thesis is easier if the aspects that RELEX can control (i.e., forecast and safety stock) are separated from lead time, batch size, order interval, and customer demand. Therefore, these two aspects are separated in this thesis when the findings are discussed.

3.7 Analysis

The purpose of the analysis phase was to utilize the results of the simulation phase and then determine how different product location parameters affect sensitivity to forecast error. This analysis section describes how stock levels and availability depend on different safety stocks, forecast errors, and product location parameters.

The analysis phase included the following four steps:

1. Analysis of stock and availability
    - Estimation of the stock level that is needed to reach a chosen availability level (95% or 99%)
2. Regression level 1
    - Calculation of the sensitivity of each product location to forecast error
3. Regression level 2
    - Analysis of how different product location parameters affect sensitivity to forecast error
4. Diagnosis

Figure 4 illustrates how data were transferred between the three first steps in the list above. The purpose of performing an *analysis of stock and availability* was to simplify supply chain performance into a single variable, *added stock*, to make later analyses easier. *Added stock* refers to how much more stock is required to compensate for forecast error. *Regression level 1* aimed to simplify sensitivity to forecast errors into two variables: *sensitivity to bias* and *sensitivity to random deviation*. *Regression level 2* created a model to describe how product parameters affect sensitivity to forecast biases and random deviations. Additionally, the created model was *diagnosed* to determine whether it fulfilled the assumptions of OLS, the chosen regression method. The regression method will be discussed in Section 3.7.3 (Regression level 2).

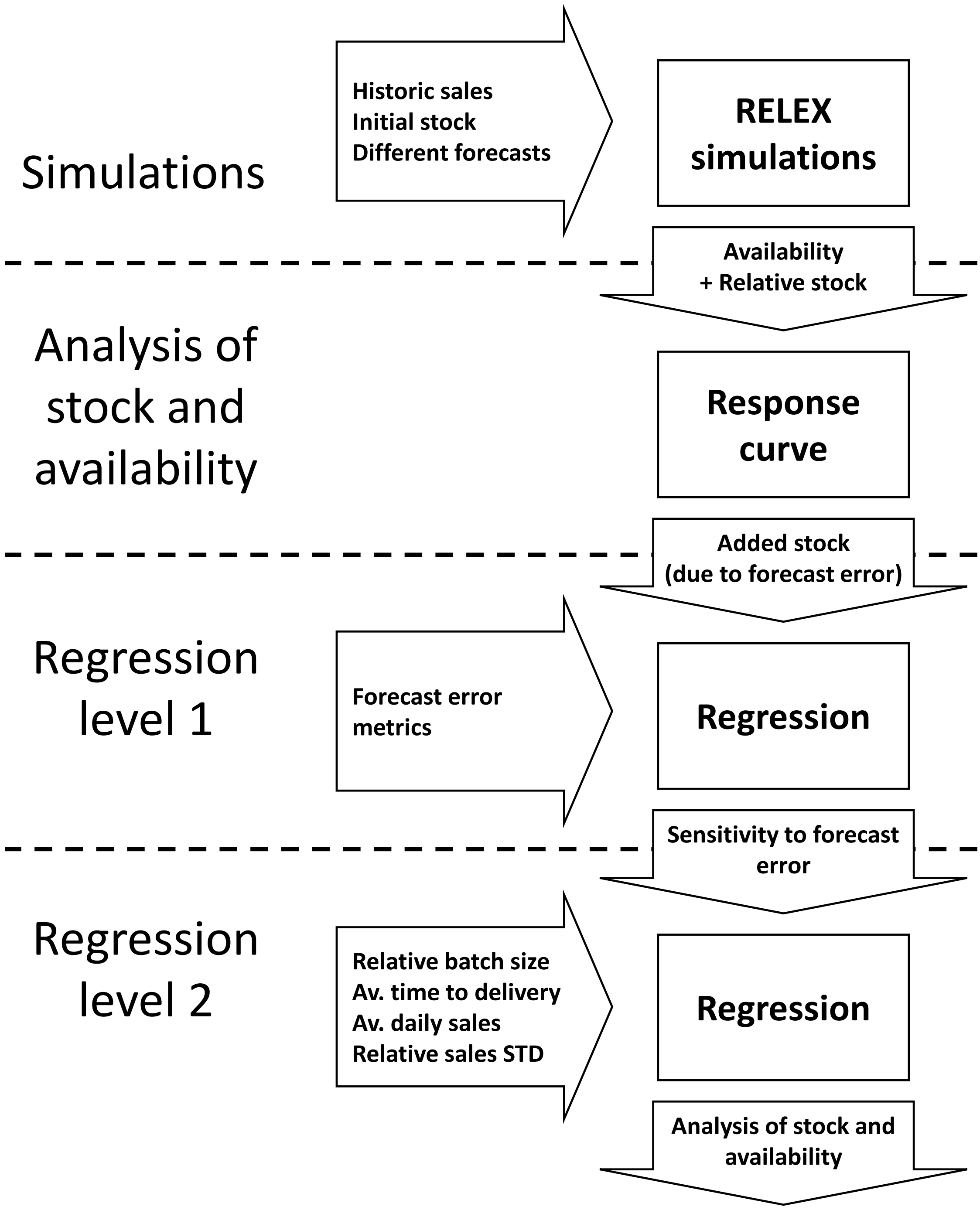

Figure 4. Analysis process

3.7.1 Analysis of stock and availability

The simulation results determine the interdependence between availability and stock. This interdependence varies to some extent between product locations, but the general structure is similar. The results of this thesis indicate that an increase in availability leads to an increase in stock. Figure 5 illustrates this interdependence. The increase in stock is higher when the availability is already high, but it is impossible to find a simple function, such as a logarithmic or quadratic function, to describe this relationship. These findings are similar to those obtained by Johnston, Taylor, & Oliveria (1988).

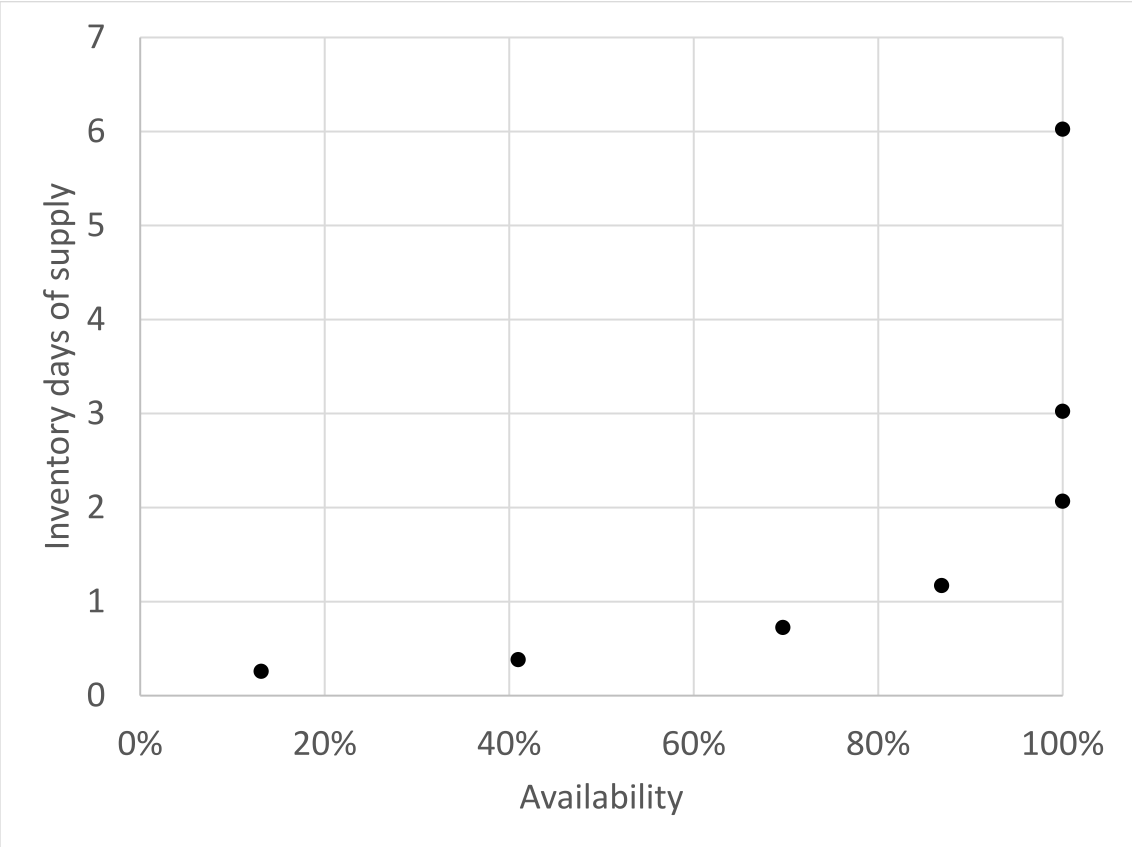

Figure 5. Example of the interdependence between availability and stock

One of the key purposes of the simulations was to calculate how deliberate errors in a perfect forecast affect the interdependence between availability and stock (i.e., the shape of the curve above) (Figure 5). Figure 6 illustrates the interdependence curves for a perfect forecast and for a forecast with a deliberate error. This approach to analyzing supply chain performance is in line with Hausman (2004).

In order to numerically analyze the phenomenon, the difference between the curves must be quantified. Figure 6 illustrates how much stock must be added in order to maintain a 95% availability level when a deliberate forecast error is introduced.

Different availability levels should produce similar results, but for the sake of completeness, two commonly used availability levels, 95% and 99%, were used in this study and the results were compared. This approach is similar to that used by Fildes & Kingsman (2011).

As availability levels are outcomes of simulations rather than inputs, it is impossible to obtain simulation results for exactly 95% or 99% availability. The required added stock for these availabilities can be interpolated by using simulation results from closely located data points. Interpolation based on the closest data points was chosen instead of modeling the whole curve, as the curve is difficult and computationally expensive to model (Johnston et al., 1988).

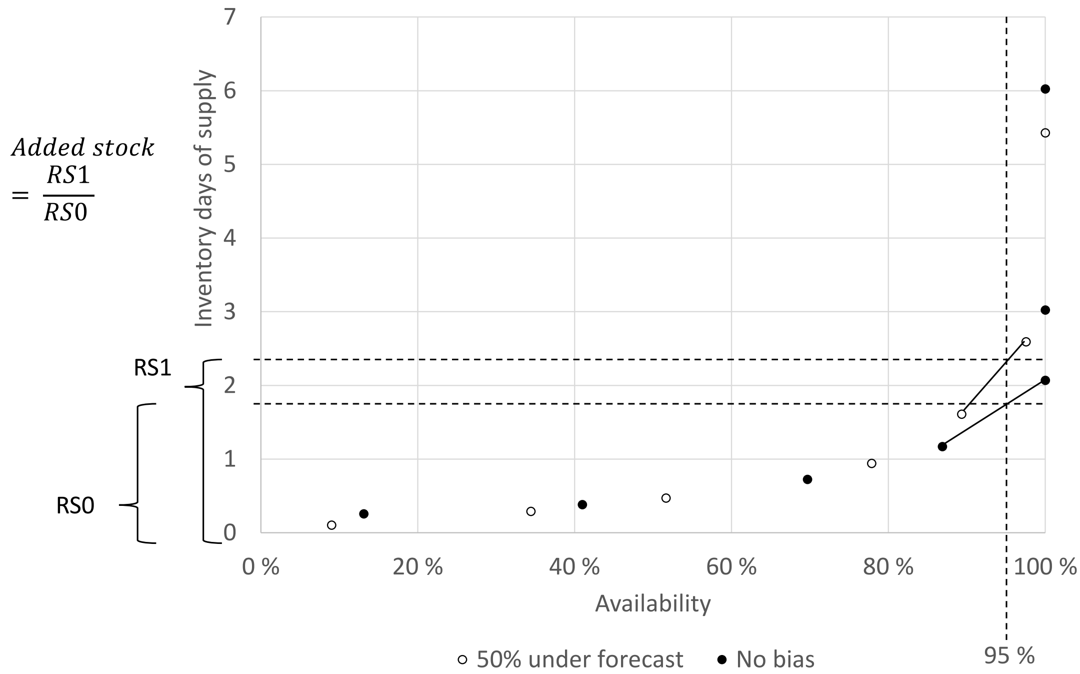

Figure 6. Effect of forecast error on the response curve

3.7.2 Regression level 1

As discussed before, the purpose of the first level of the regression was to simplify sensitivity to forecast errors into two variables: *sensitivity to bias* and *sensitivity to random deviation*.

Two linear regression models were created for each product location: one for bias and another for random deviation. The independent variable in the models was forecast bias or random deviation, and the dependent variable was added stock. Figure 7 and Figure 8 illustrate the interdependence between added stock and bias/random deviation.

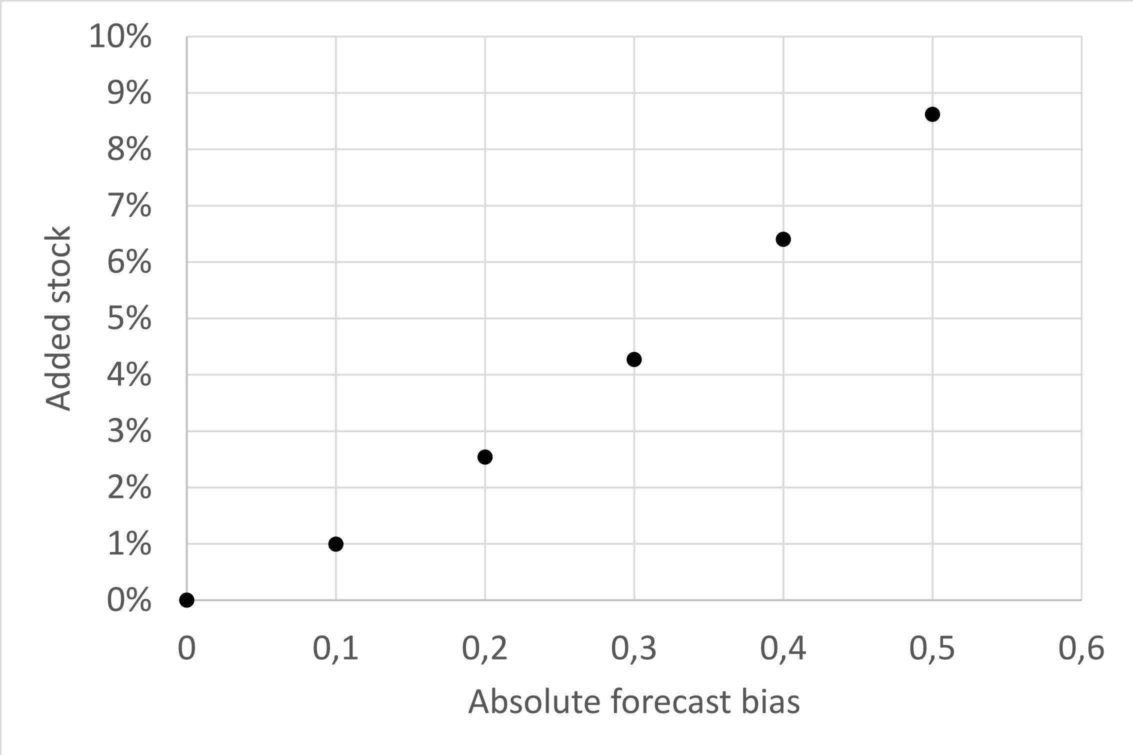

Figure 7. Effect of forecast bias on added stock

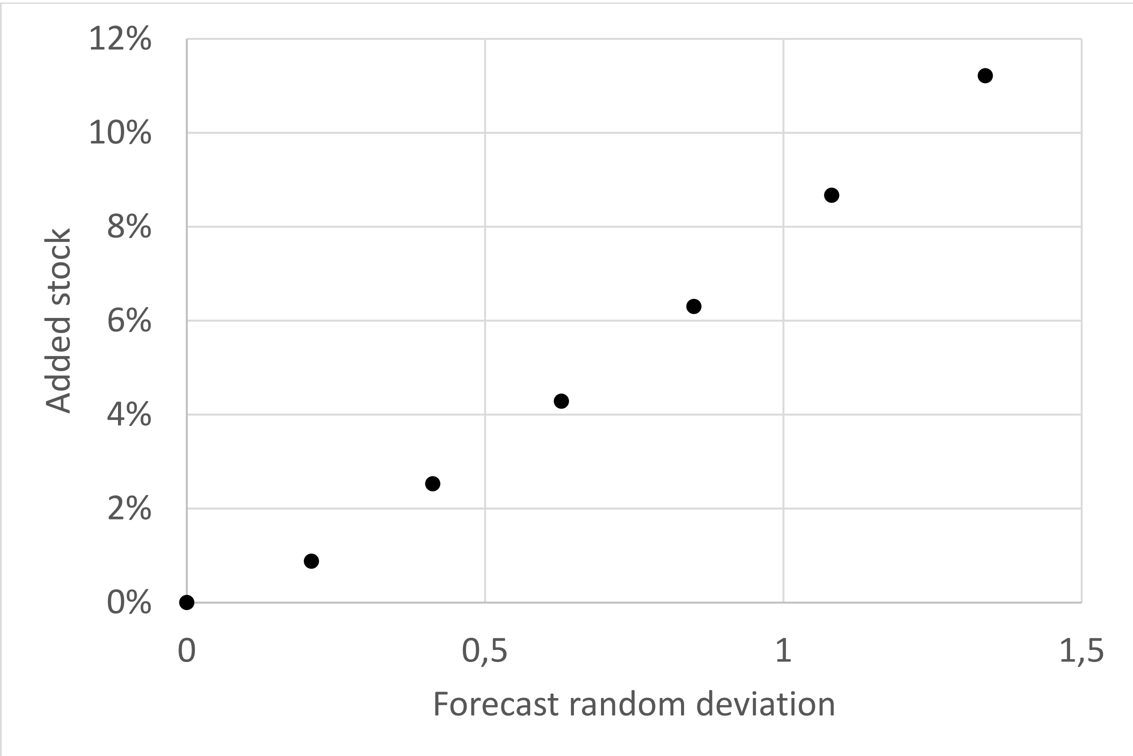

Figure 8. Effect of random deviation of forecasts on added stock

In mathematical terms, the relationships between added stock and bias/random deviation are defined as follows:

$$y = a + \sum k_i x_i + e \tag{10}$$

where y *= added stock*
a *= constant*
x *= bias or random deviation*
k *= regression coefficient*
e *= error term*

The two created models produced coefficients for each product location. The obtained coefficients can be interpreted as the product location's sensitivity to forecast bias and random deviation.

3.7.3 Regression level 2

In the first level of the regression, sensitivity to forecast error was calculated separately for each product location. The second level of the regression uses product location parameters to explain why different product locations have different sensitivities.

As was the case for the first level, there are two linear regression models for the second level: one for bias and one for random deviation. The product location's sensitivity to forecast error was the dependent variable, and product parameters, such as batch size, were the independent variables.

In mathematical terms, the relationships between sensitivity to forecast bias/random deviation and product location parameters are defined as follows:

$$y = a + \sum k_i x_i + e \tag{11}$$

where y *= sensitivity to bias/random deviation*
a *= constant*
x_i *= parameter value*
k_i *= regression coefficient*
e *= error term*

In Formula (11), there are five values for k and x, one for each product location parameter. In other words, i can refer to relative batch size, lead time, order interval, average daily sales, or relative sales STD.

4 Model development

The purpose of this thesis was to create a statistical model to explain product location's sensitivity to forecast error. When created, a model should describe reality as closely as possible. OLS is a statistical method that is commonly used to explain observations about a phenomenon. In order to use OLS, some key requirements should be met, including multicollinearity, linearity, and normality (Mendenhall & Sincich, 2011). There should be no multicollinearity (i.e., independent variables should not correlate with each other). The relationships between independent and dependent variables should be linear. The residuals (i.e., deviations between the model and observations) should be normally distributed. These three key criteria are discussed in the following sections.

4.1 Multicollinearity

In order to obtain a good model, the independent variables should be linearly independent (i.e., the multicollinearity of different variables should be minimized). Table 2 illustrates the correlations between different variables, including both the dependent variables (i.e., sensitivity to random deviation and sensitivity to bias) as well as the independent variables (i.e., relative batch size, lead time, order interval, average daily sales, and relative sales STD). Correlations between the independent variables can be problematic and may violate the assumptions of OLS, while correlations between dependent variables are allowed and even expected. For clarity, the two dependent variables are highlighted in grey below.

Table 2. Correlation matrix

Variable	1)	2)	3)	4)	5)	6)	7)
1) Sensitivity to random deviation	1						
2) Sensitivity to bias	0.81	1					
3) Relative batch size	-0.36	-0.34	1				
4) Lead time	0.12	0.33	-0.3	1			
5) Order interval	0.11	0.26	-0.2	0.87	1		
6) Av. daily sales	0.81	0.64	-0.35	0.07	0.11	1	
7) Relative sales STD	-0.47	-0.46	0.75	-0.33	-0.27	-0.48	1

Unfortunately, there are some rather high correlations. The highest correlation appears to be between lead time and order interval (highlighted in red), which is a potential problem. Also, the correlation between sensitivity to bias and sensitivity to random deviation and that between average daily sales and sensitivity to random deviation appear to be high (highlighted in green). However, the latter two are not problematic as sensitivity to random deviation is a dependent variable.

In order to assess the severity of multicollinearity, an additional indicator, variance inflation factor (VIF), was applied. The left column of Table 3 presents the VIFs for all the independent variables. The VIFs for lead time and order interval appear to be relatively high (highlighted in red). It was realized that the situation could be improved by combining lead time and order interval into one variable, average time to delivery, using Formula (3):

$$Av.\,time\ to\ delivery = Lead\ time + \frac{order\ interval}{2} \tag{3}$$

Average time to delivery refers to the mean time until the next possible delivery. The right column of Table 3 presents the VIFs after this transformation. As the right column shows, the situation has significantly improved (highlighted in green).

Table 3. VIFs

With all the values separate		With two values combined	
Variable	VIF	Variable	VIF
Relative batch size	2.375215	Relative batch size	2.316457
Lead time	4.440452	Av. time to delivery	1.129104
Order interval	4.193709	Av. daily sales	1.318292
Av. daily sales	1.344835	Relative sales STD	2.750745
Relative sales STD	2.750897		

4.2 Linearity

In order to satisfy the assumptions of OLS, the relationships between independent and dependent variables should be linear. As Figure 9 shows, the relationship between sensitivity and relative batch size was not linear. This problem could be corrected by applying a suitable transformation to the independent variable: in this case, relative batch size. A potential solution is to use the mathematical inverse of relative batch size—1/(relative batch size)—which can be interpreted as the number of batches sold per day. Figure 10 shows that the linearity requirement could be better achieved using *batches per day* instead of relative batch size, and therefore, the former proved to be a better alternative to the independent variable in question.

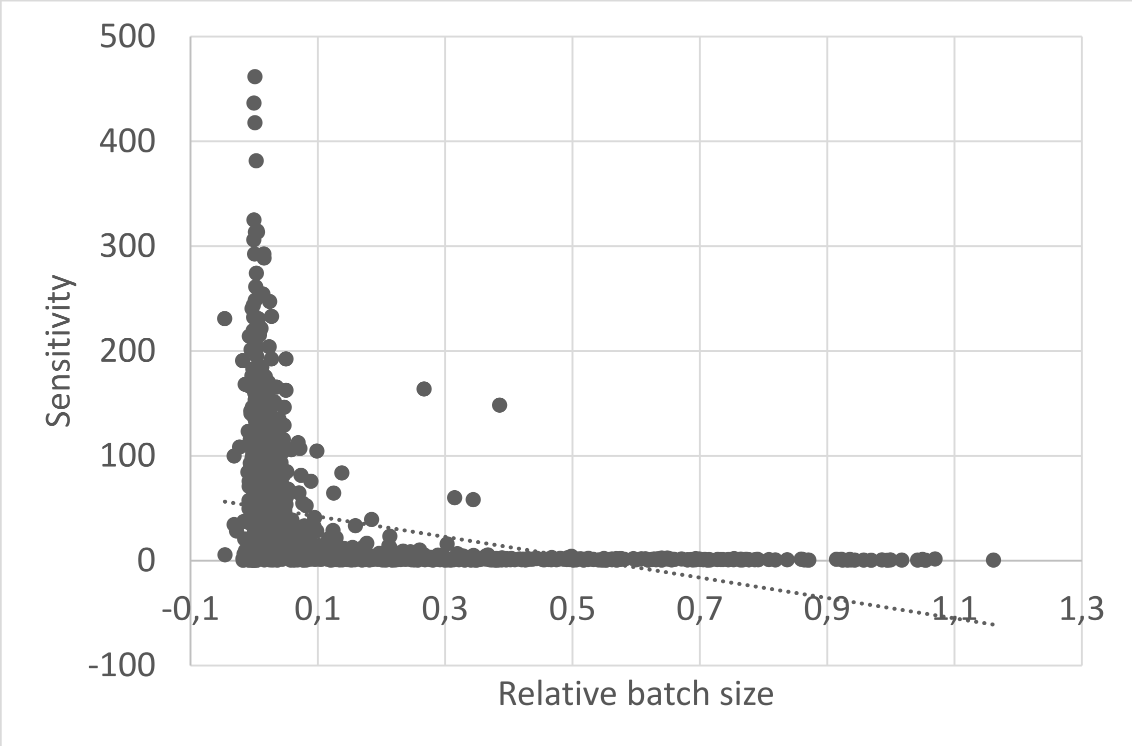

Figure 9. Relationship between sensitivity to random forecast deviation and relative batch size

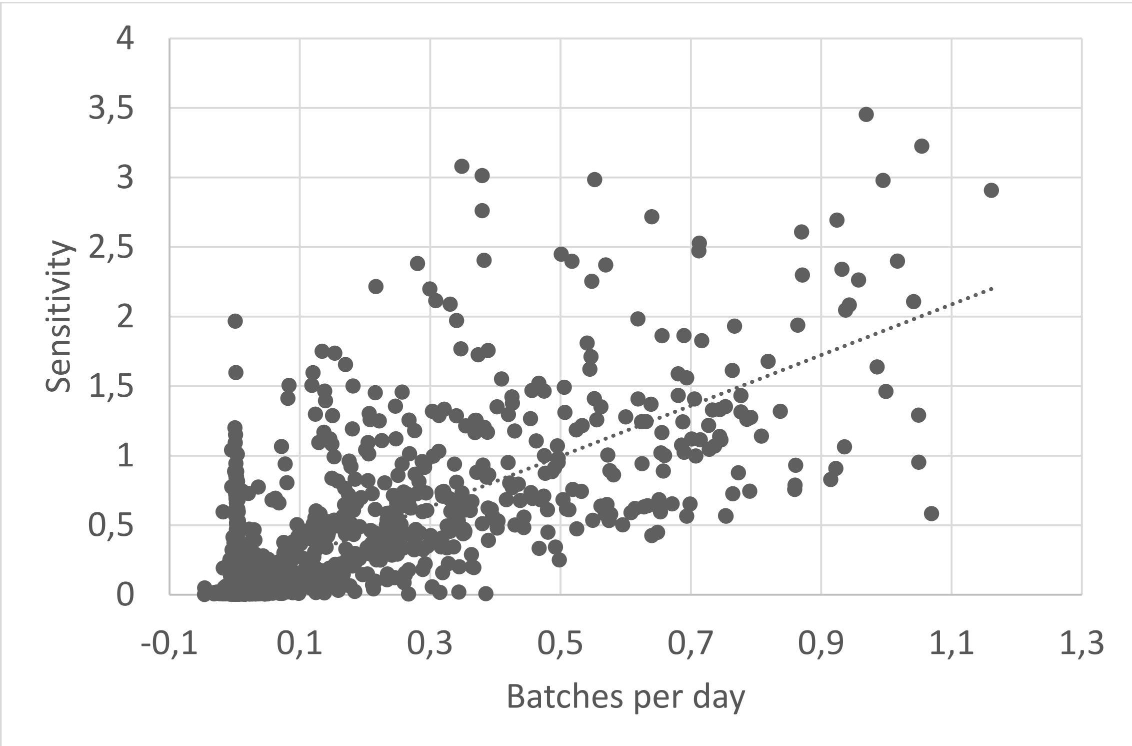

Figure 10. Relationship between sensitivity to random forecast deviation and batches per day

4.3 Forecast error

Using the parameters defined in the previous sections, it was possible to build a linear model with *sensitivity to forecast error* as an independent variable and *relative batch size, average time to delivery, average daily sales* and *relative sales STD* as dependent variables. The key characteristics of this linear model are presented in Table 4, which reveals some interesting observations. As the significance column shows, three of the variables proved to be statistically significant (highlighted in green). First, the factor batches sold per day seems to have a positive effect on sensitivity to forecast error. In other words, large batch sizes make forecast accuracy less important. Second, average time to delivery and average daily sales have a positive effect on the importance of forecast accuracy, that is, large values make forecast accuracy more important. Relative sales STD was found to be statistically insignificant.

Table 4. Linear model of what affects sensitivity to forecast error

Variable	Estimate	Pr(>|t|)	Significance
(Intercept)	0.00453	0.83896	
Batches sold per day	0.21617	0.00000	***
Av. time to delivery	0.02568	0.00000	***
Av. daily sales	0.02709	0.00000	***
Relative sales STD	0.00222	0.71618	

* significant at p<0.05; ** significant at p<0.01; *** significant at p<0.001

4.4 Forecast bias and random deviation

As suggested by Barman, Tersine, & Burch (1990), Kerkkänen, Korpela, & Huiskonen (2009), and Sanders & Graman (2009), forecast *bias* and *random deviation* might have different effects on supply chain performance. For this reason, there was a need to create separate models for bias and random deviation to test whether there is a difference between the two. However, Table 5 shows that bias and random deviation behaved similarly in the created models. In particular, *batches sold per day* and *average daily sales* had a positive effect on both sensitivity to bias and sensitivity to random deviation. However, *average time to delivery* was only significant for sensitivity to bias.

Table 5. Comparison between forecast bias and random deviation

Sensitivity to bias			Sensitivity to random deviation		
Variable	Estimate		Variable	Estimate	
(Intercept)	-0.04130	***	(Intercept)	0.03793	***
Batches sold per day	0.08741	***	Batches sold per day	0.11997	***
Av. time to delivery	0.02673	***	Av. time to delivery	0.00185	
Av. daily sales	0.00968	***	Av. daily sales	0.01615	***
Relative sales STD	-0.00981	**	Relative sales STD	-0.00753	**

* significant at $p<0.05$; ** significant at $p<0.01$; *** significant at $p<0.001$

4.5 Positive and negative biases

According to prior literature, there might be a need to model positive and negative biases separately as they might behave differently (Barman et al., 1990). However, Table 6 shows that positive and negative biases had very similar effects. Therefore, it was sufficient to analyze only the magnitude of bias.

Table 6. Comparison of positive and negative bias

Sensitivity to pos. bias			Sensitivity to neg. bias		
Variable	Estimate		Variable	Estimate	
(Intercept)	-0.04117	***	(Intercept)	-0.04473	**
Batches sold per day	0.06950	***	Batches sold per day	0.10767	***
Av. time to delivery	0.03540	***	Av. time to delivery	0.01757	***
Av. daily sales	0.00575	***	Av. daily sales	0.01320	***
Relative sales STD	-0.01636	***	Relative sales STD	-0.00138	

* significant at p<0.05; ** significant at p<0.01; *** significant at p<0.001

4.6 Different availability levels

It was crucial to test whether the models produced different results for different availabilities. Availability levels of 95% and 99% are commonly used in related literature (Fildes & Kingsman, 2011; Syntetos et al., 2010). Table 7 illustrates how the independent variables affect sensitivity to forecast error in cases with 95% and 99% availability. The results proved similar in regards to batches sold per day, average daily sales, and relative sales STD. However, average time to delivery was not significant with 99% availability.

Table 7. Comparison of 95% and 99% availability levels

95% availability			99% availability		
Variable	Estimate		Variable	Estimate	
(Intercept)	0.00453		(Intercept)	0.12181	***
Relative batch size	0.21617	***	Relative batch size	0.20807	***
Av. time to delivery	0.02568	***	Av. time to delivery	-0.00453	
Av. daily sales	0.02709	***	Av. daily sales	0.03722	***
Relative sales STD	0.00222		Relative sales STD	-0.00671	

* significant at p<0.05; ** significant at p<0.01; *** significant at p<0.001

4.7 Distribution of residuals

The OLS regression methods used in this thesis assume normally distributed residuals (i.e., deviations between the model and observations). Figure 11 illustrates the obtained residuals for one of the developed models. The residuals for other models were very similar. The residuals were relatively close to the normal distribution. Figure 12 illustrates the difference between normal distribution and the distribution of the obtained residuals. According to Sheather (2009), the plot should be close to a straight line for the data to be consistent with that of a normal distribution. Figure 12 seems to be close to a straight line.

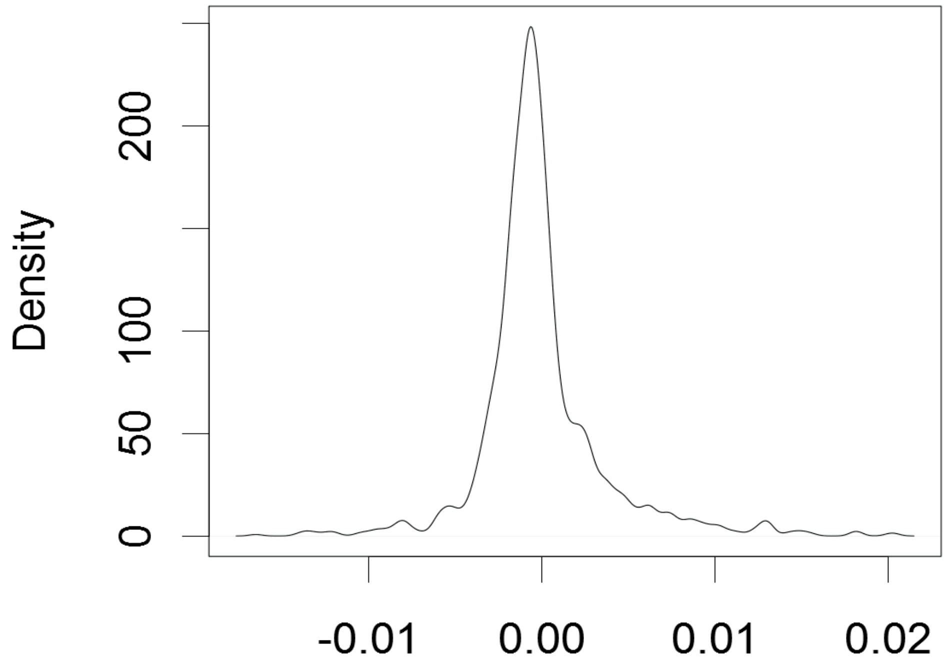

Figure 11. Distribution of the residuals

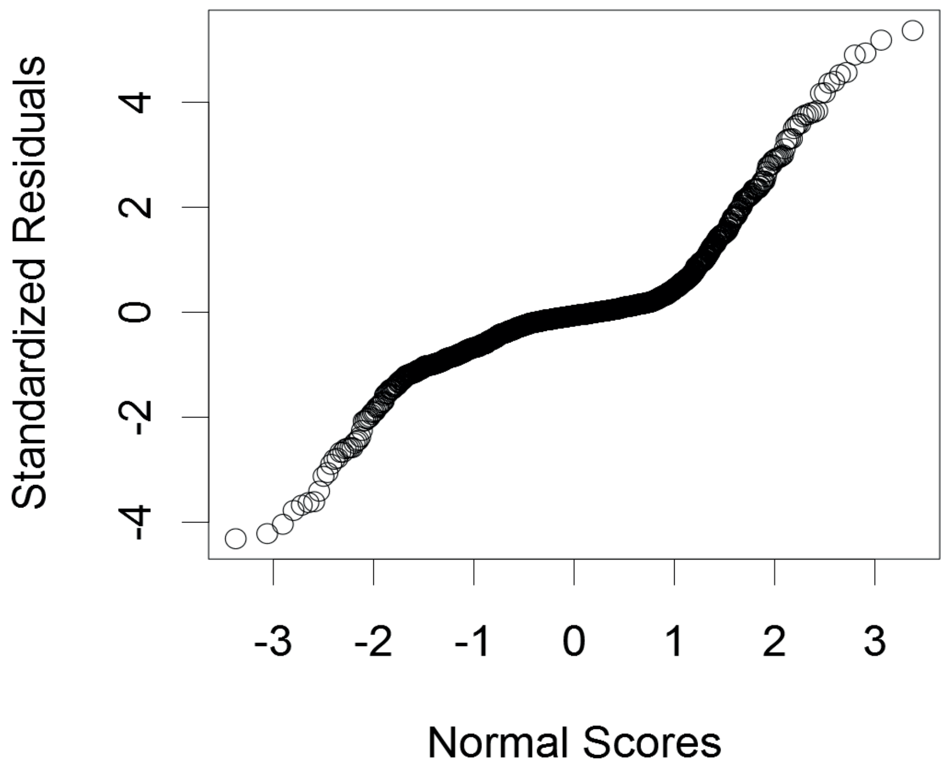

Figure 12. Quantile-quantile plot illustrating the difference between the observed and assumed distributions of residuals

5 Results

5.1 Effect of forecast error on added stock

Based on simulations using real sales data from a major retail customer, this thesis clarified how supply chain performance is affected by forecast errors. The results quantify how much additional stock is needed to compensate for forecast error. The effect of forecast errors was found to be linear for both bias and random deviation. The results were obtained by expressing bias and random deviation as described in Section 2 (Literature review), i.e., RCFE and RFSD, respectively.

5.2 Effect of product location parameters on sensitivity to forecast error

Chapter 4 (Model development) discussed the statistical significance of each of the independent variables. From a business viewpoint, the key issue is to quantify the importance of each of the variables regarding sensitivity to forecast error. Figure 13 illustrates the importance of *average daily sales, relative batch size, average time to delivery*, and *relative sales STD*. Importance was calculated by comparing the R^2 values of different models. The R^2 values describe how well each model explains variations in the dependent variable.

First, a model was created with sensitivity to forecast error as a dependent variable and all four product location parameters as independent variables. Then, another model was created with one of the independent variables removed. The R^2 values of the models were compared, and the difference was interpreted as the importance of the removed variable. The same process was repeated for each of the independent variables. Finally, the importance values were scaled to add up to 100%. The differences in importance proved to be very clear, indicating a logical basis for prioritization.

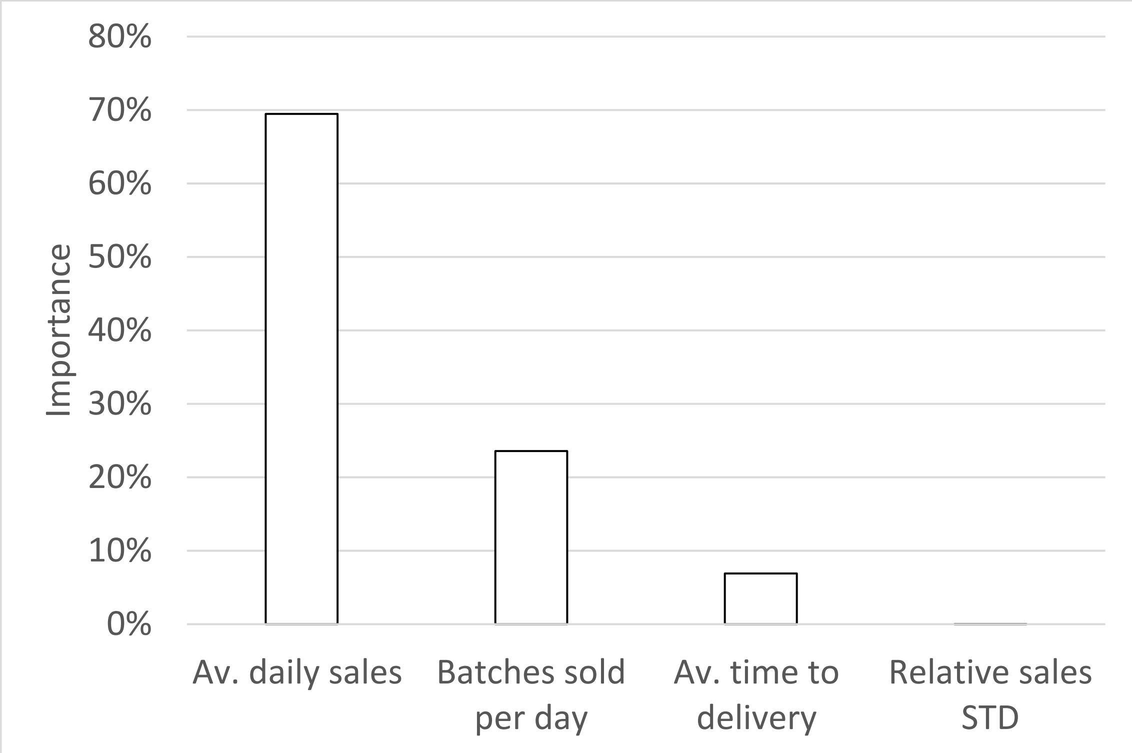

Figure 13. Importance of each of the independent variables based on their R^2 values

5.2.1 Average daily sales

Average daily sales proved to be the most significant factor affecting the importance of forecast accuracy. Sensitivity to forecast accuracy declined sharply with a decrease in average sales, as depicted in Figure 14. In this figure, sensitivity to forecast error is defined using relative changes in stock, that is, the percentage of additional stock needed to achieve the same availability. From a business viewpoint, absolute figures are more important than relative ones. Consequently, both stock and availability are very important for products with higher sales. Figure 15 repeats the information presented in Figure 14 but scales the values by average inventory. Thus, Figure 15 better represents the business viewpoint. Figure 15 has absolute sensitivity on the y-axis regarding business importance, whereas Figure 14 has relative sensitivity.

By comparing Figure 14 and Figure 15, it is clear that *average sales* of a product location is more important than what initially appeared. This finding highlights the need to concentrate on high-selling products when efforts to improve forecast accuracy are considered. Table 8 presents additional information related to Figure 14 and Figure 15, illustrating the number of product locations in different average daily sales categories.

Table 8. Number of product locations in different average daily sales categories

Average daily sales (pcs)	Product locations	% of product locations
μ < 1/30	22	2 %
1/30 < μ < 1/7	315	23 %
1/7 < μ < 1	481	35 %
1 < μ < 5	156	11 %
5 < μ < 10	196	14 %
μ > 10	210	15 %

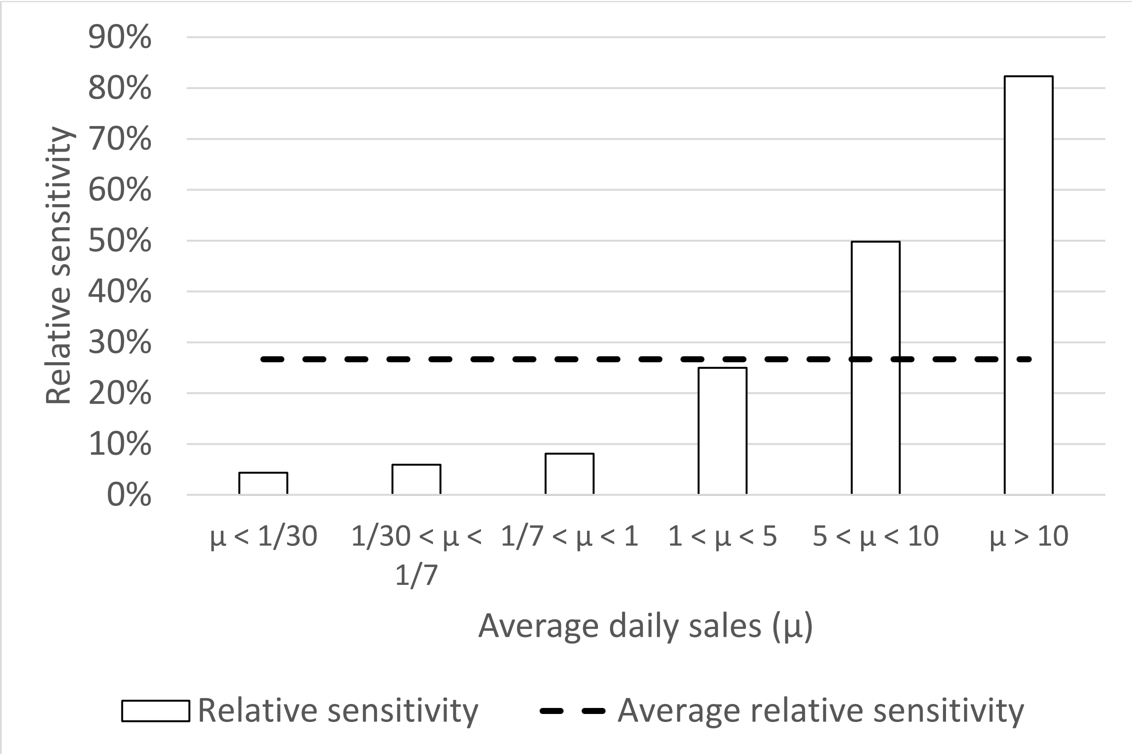

Figure 14. Effect of sales frequency on relative sensitivity to forecast error based on the sample data. μ = average unit sales per day. For example, μ < 1/30 means less than one unit sold per month, and μ > 5 means more than five units sold per day. Relative sensitivity is the sensitivity to forecast error (i.e., the percentage of stock that is added with one unit of change in forecast error).

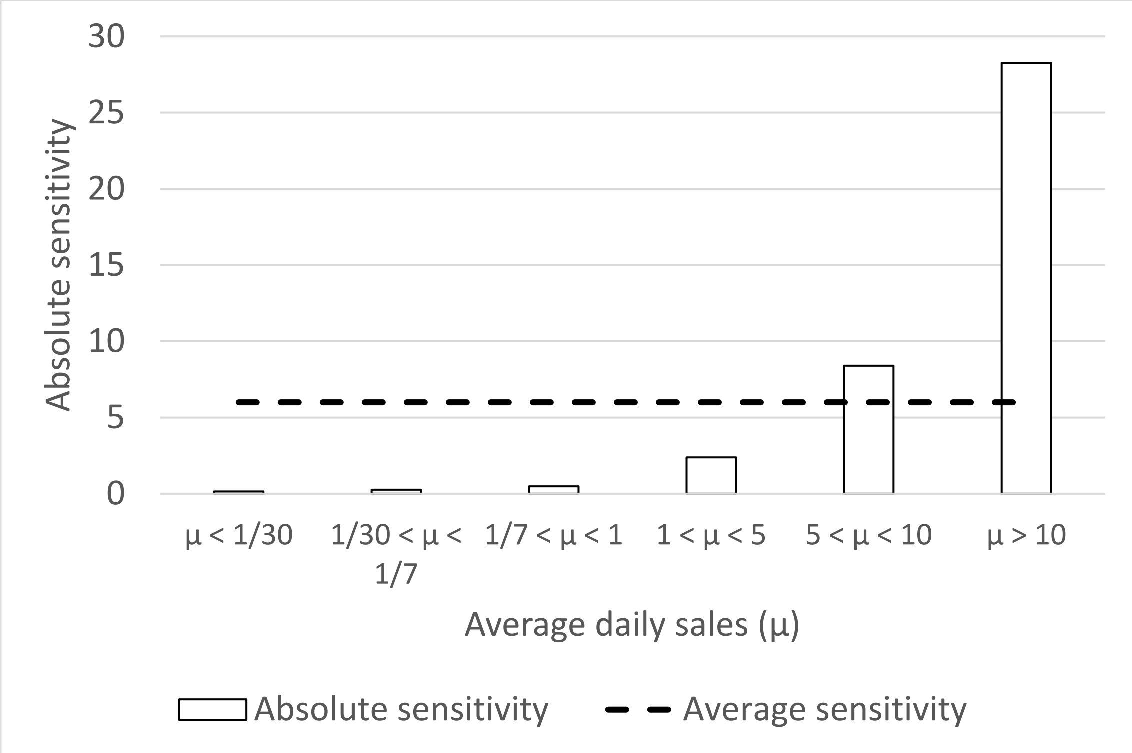

Figure 15. Effect of sales frequency on absolute sensitivity to forecast error based on the sample data. Absolute sensitivity is the sensitivity to forecast error (i.e., how many units of stock are added with one unit of change in forecast error).

A low-selling product may have a large relative forecast error; however, this is not necessarily important for business. One specific problem with forecasting low-selling products is the discrete nature of product sales; products can only be sold in whole numbers, but forecasts can produce decimal values. For example, a product selling once a week on average could have a forecast of 0.14 per day. This could produce fairly accurate orders and low stock levels but still be associated with a fairly large relative forecast error.

5.2.2 Relative batch size

The second most important factor affecting sensitivity to forecast error was *relative batch size*, which can also be expressed as *batches sold per day*. Figure 16 illustrates the effect of this factor. Large batch sizes result in higher average stock, causing forecast errors to become less critical. With large relative batch sizes, forecast errors only matter at the end of the replenishment cycle. Increasing relative batch size leads to longer replenishment cycles, resulting in fewer days on which stocks are low and forecast accuracy matters. This could also explain why sensitivity to forecast error and relative batch size did not follow a linear relationship, but rather seemed to be inversely related. This makes sense; for example, doubling the batch size would halve the number of days on which forecast accuracy is important.

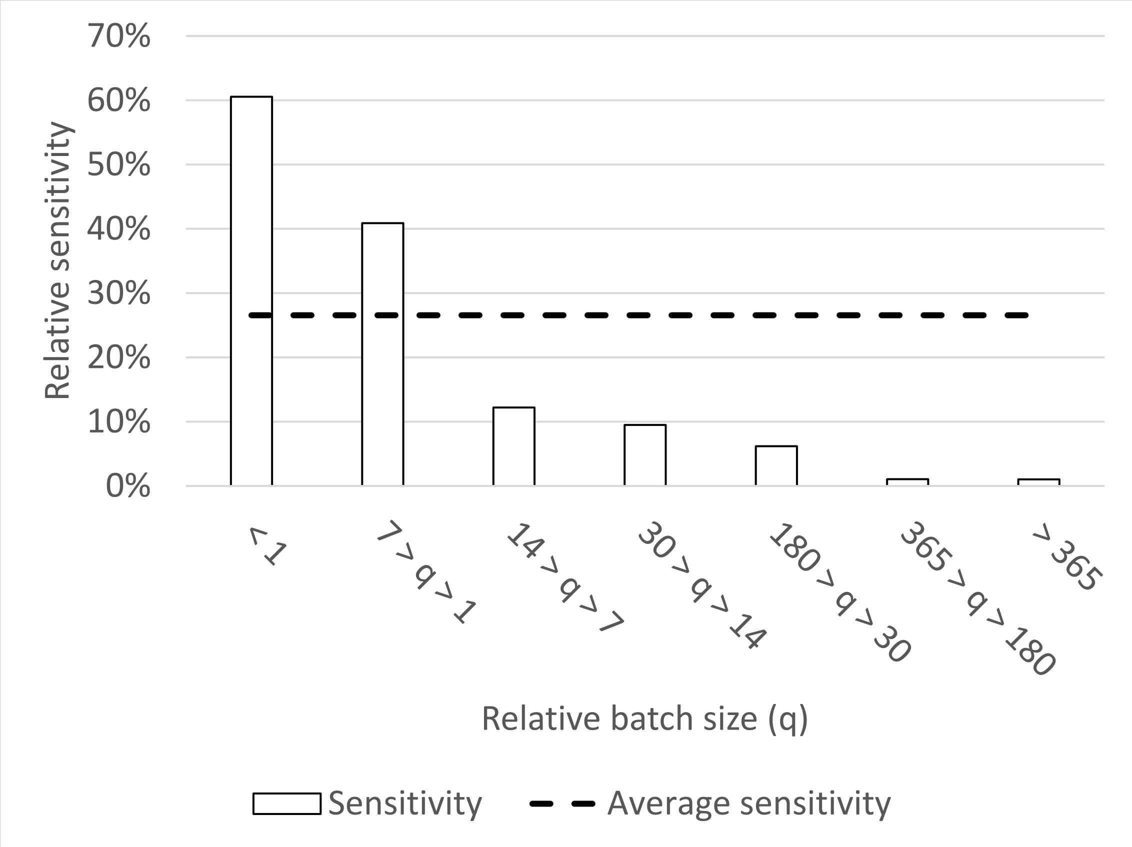

Figure 16. Effect of relative batch size on relative sensitivity to forecast error based on the sample data. q = relative batch size. For example, q < 1 means that one batch does not cover the sales of an average day. Relative sensitivity is the sensitivity to forecast error (i.e., the percentage of stock that is added with one unit of change in forecast error).

5.2.3 Average time to delivery

As shown in Figure 13, *average time to delivery* had some effect on the importance of forecast accuracy, but the effect was relatively small. This is most likely due to short lead times and review periods, as is typical in the retail industry. In the analyzed business data, most lead times were only one day long. Very few products had lead times of one week, and none had lead times longer than one week. Therefore, it is logical that average time to delivery was not critical.

Average time to delivery was found to be a statistically significant factor for sensitivity to bias but not for sensitivity to random deviation of forecasts. This is understandable; in the case of bias, forecast errors accumulate over time and cause stock-outs or excess stock if the uncertainty period is long, while in the case of random deviation, forecast errors tend to cancel each other out over periods of time. Thus, average time to delivery did not prove significant.

As discussed in Section 4.6 (Different availability levels), average time to delivery was a statistically significant factor with 95% availability but not with 99% availability. This can be explained by the relatively small number of products with high average time to delivery and high safety stocks caused by the high level of availability. The high safety stocks were likely capable of absorbing the relatively minor effects of the few products with longer times to delivery.

5.2.4 Relative sales STD

Relative sales STD was analyzed since it was discussed in the scientific literature. However, the results of this study show that this variable is not important from the business perspective (see Figure 13).

5.2.5 The impact of product location parameter variations

Table 9 illustrates the impact of variations in the values of product location parameters if the random deviation of forecasts decreases from 3 to 2 when measured using WAPE. Table 9 compares the changes in inventory when daily sales are increased five-fold, batch size is decreased to one-fifth of the original size, and time to delivery is increased seven-fold. In other words, the table includes two different values for each of the three product location parameters.

Table 9. Example of the impact of product location parameters

	Batch size	Time to delivery	Daily sales	Expected change in inventory
Case 1	10	1	2	7.46 %
Case 2	10	1	10	16.36 %
Case 3	2	1	2	12.18 %
Case 4	10	7	2	8.11 %

Table 9 shows that the most important parameter is daily sales, followed by batch size, as these factors showed the largest changes in inventory.

5.2.6 Summary of the impact of all product location parameters

Figure 17 summarizes the impact of all product location parameters by illustrating the importance of the most important product locations. The figure shows the percentage of improvement that can be obtained by taking into account the top 1%, 2%, 5%, 10%, 20% or 50% of product locations. The grey bars illustrate the impact of the theoretical maximum (i.e., if we knew the exact absolute importance of each product location). The white bars illustrate the impact of using the model developed in this thesis to choose which product locations are taken into account.

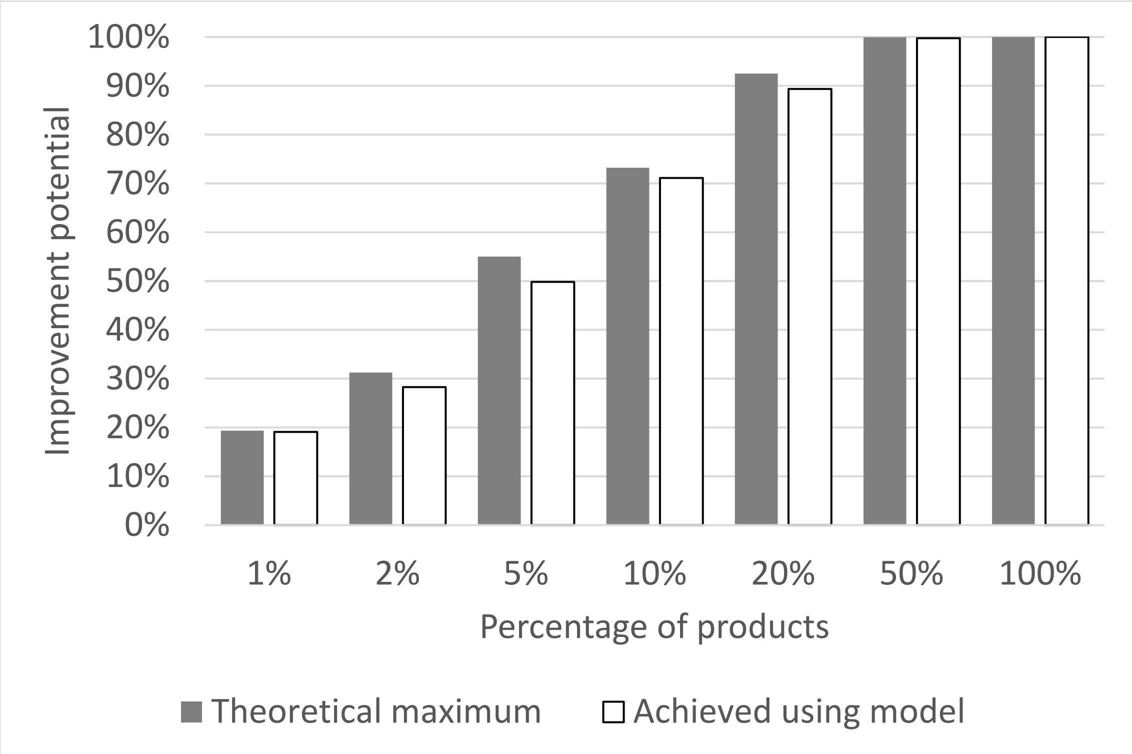

Figure 17. The importance of the highest-selling product locations

6 Discussion and conclusions

This thesis aimed to clarify the situations in which forecast accuracy is important. In order to achieve this goal, real sales and product data from a major retail customer of RELEX were used to simulate the effects of different deliberately modified forecasts.

6.1 Scientific implications

The scientific literature identifies several product location parameters that affect the importance of forecast accuracy, including *batch size* (Ritzman & King, 1993; Fildes & Kingsman, 2011), *lead time* (Ott et al., 2013), *sales volume* (Hoover, 2009; Ott et al. 2013; Kerkkänen et al. 2009), and *unevenness of demand* (Ritzman & King, 1993). However, the scientific literature fails to prioritize these parameters. This thesis contributes to scientific knowledge by suggesting a clear priority for the business importance of the key product location parameters in the retail context.

The results of this study indicate that a product's *average daily sales* is the most significant factor affecting the importance of forecast accuracy. *Relative batch size* proved to be the second most important product location parameter. *Average time to delivery* was the third most important parameter, and *relative sales STD* was the fourth. It is worth noting that average time to delivery was significant only in the case of forecast bias, something that was not found in literature.

The results of this study showed a positive correlation between *availability* and *inventory days of supply*, but this relationship did not follow a simple mathematical model. These observations are in line with those obtained by other researchers (Fildes & Kingsman, 2011; Johnston et al., 1988).

In this study, the effect of forecast errors on supply chain performance was found to be linear. This covered errors due to both bias and random deviation, as measured by RCFE and RFSD. The literature seems to have mixed results on this topic; some articles reported a linear relationship between supply chain performance and forecast error, while others found the relationship to be exponential (Fildes & Kingsman, 2011; Sanders & Graman, 2009). These mixed results could be explained by the fact that different studies used different metrics in their analyses. An additional reason might be that this research used real sales data, while most researchers used simulated data (Biggs & Campion, 1982; Xie et al., 2004).

The results of this study showed no significant difference between the effects of positive and negative biases, even though the literature suggests that there might be differences (Barman et al., 1990; Biggs & Campion, 1982; Ritzman & King, 1993). This can be logically explained by the use of a single variable, *supply chain performance*, instead of two (or more) different variables, *stock* and *availability*, as is often done in the scientific literature. Successful simplification of supply chain performance into one variable in the retail context is in itself a contribution to the scientific knowledge.

6.2 Practical implications

6.2.1 Implications for retail business

This thesis aimed to clarify the situations in which manual efforts to improve forecast accuracy are valuable for retail companies. The results of this study suggest clear principles for prioritizing product locations to achieve the maximum benefit from manual corrections. In addition, clear prioritization principles may be useful when considering investments in more accurate methods, such as weather-based forecasting or analysis of competitors' actions. These advanced forecasting methods require reliable data and are costly. Therefore, the business benefits of such investments need to be proven.

The findings of this thesis highlight the importance of *average daily sales* and *relative batch size* over other product location parameters. Therefore, retailers should consider focusing their efforts primarily on improving forecast accuracy for high-selling products and, to some extent, for products with small batch sizes.

Average time to delivery and *relative sales STD* were both found to be statistically significant at least in some cases. However, as pointed out by Fildes & Kingsman (2011), statistical significance is not necessarily associated with economic importance of the same magnitude. From the business perspective, *sales volume of a product location* proved to be the most important parameter, and *relative batch size* had some importance. Other product location parameters are relatively unimportant when compared to these two, although *average time to delivery* might play a minor role. The role of *relative sales STD* proved so marginal that retail managers could consider ignoring it.

The clear prioritization principles discussed above may need justification before implementation within retail companies and their suppliers. In particular, there is a need to explain why average time to delivery and relative sales STD play only minor roles. The minor role of average time to delivery can be explained by relatively short lead times and review periods for products in the case company, as is typical in retail businesses. In other settings, such as warehouses, there might be a wider variety in lead times, potentially increasing the importance of lead times. A logical explanation for why relative sales STD proved unimportant in this thesis is that this viewpoint is already covered by the most important product location parameter (i.e., average daily sales). Product locations with high average daily sales tend to also have low relative sales STD, and therefore it is sufficient to analyze only the more important factor of these two.

6.2.2 Implications for RELEX

The results of this study suggest clear principles for retail companies to prioritize product locations to achieve the maximum benefit from manual corrections. RELEX may benefit from the results by offering this prioritization functionality as part of their product.

This thesis managed to quantify how much additional stock is needed to compensate for forecast error. The results also revealed the impact of different product location parameters on the business importance of forecast accuracy. The obtained results may help RELEX's sales efforts as it will be easier to show how much potential new customers could save with better forecasts. Therefore, it will be easier to prove that investing in RELEX's forecasting solutions is profitable for retail customers. In addition, knowing the customer's potential savings could help RELEX set prices for its solutions. Furthermore, knowing when forecast accuracy is not important could help when negotiating with customers demanding unnecessary accuracy.

RELEX often deals with competing companies that offer complicated and non-transparent forecasting methods. The results of this thesis may provide research-based evidence for criticizing competitors' sales arguments if the potential improvements in accuracy can be shown to be on products for which improved forecast accuracy does not have any noteworthy business benefit.

Both RELEX and its competitors offer advanced forecasting methods, such as weather-based forecasting, which require a lot of sales data to be reliable. The results of this thesis may give arguments for selling these advanced functionalities. The results favor prioritization of highly selling products, which also benefit from advanced forecasting methods. Through this synergy, it is a logical decision for customers to invest in these advanced functionalities.

6.3 Generalizability and limitations

This study was based on sales and product data from a large European retailer, with quickly spoiling products excluded. Therefore, it can be assumed that the findings are generalizable to companies similar to the case company. Typical beneficiary companies could be relatively large retailers and chains with a significant part of their revenue coming from fast moving consumer goods with reasonably long shelf lives. Retailers with a slightly different profile might also benefit from the findings of this study to a certain extent.

There are a few limitations in this study regarding the generalizability of the findings. First, the study only analyzed one case company. Product location parameters might have significantly differing values in different companies, potentially changing the prioritization of the parameters. However, these parameters depend more on products and suppliers than retail companies, and therefore the problem might not be severe.

Second, artificially generated forecasts might not correctly simulate all the properties of real forecasts. Thus, it is possible that forecast errors occurring in reality differ from the artificially generated ones.

Another potential problem could be that the curve describing the relation between *availability* and *inventory days of supply* was not continuous in this study. Therefore, linear interpolation was required for values between the data points. Linear interpolation introduces small errors as the relationship is not totally linear. However, the distance between data points was not large, and therefore these errors should not be significant.

6.4 Further research

A natural next step is to perform similar simulations in other companies and in other sectors to increase the generalizability of the findings. For retail companies, it would be beneficial to expand this study by including products with short shelf lives. However, this would require mathematically modeling spoilage.

This study focused on retail stores. Another potential way to expand this study is to investigate how different players in the supply chain are affected by forecast errors and the same product location parameters that were studied in this master’s thesis.

It would also be interesting to study whether analyzing cumulative human experiences and views would result in different conclusions than this purely quantitative study, which was based on simulations and statistical analyses. Therefore, qualitative studies involving interviews with experienced representatives of different actors in the supply chain could be an area for further research.

References

Agrawal, N., & Smith, S. (2015). *Retail supply chain management: quantitative models and empirical studies* (Vol. 122). http://doi.org/10.1007/978-0-387-78902-6

Ayers, J. B., & Odegaard, M. A. (2007). *Retail Supply Chain Management* (Vol. 26). http://doi.org/10.1007/978-1-4899-7562-1

Barman, S., Tersine, R. J., & Burch, E. E. (1990). Performance evaluation of the LDR and the PSH with forecast errors. *Journal of Operations Management*, *9*(4), 481–499. http://doi.org/10.1016/0272-6963(90)90006-Y

Biggs, J. R., & Campion, W. M. (1982). THE EFFECT AND COST OF FORECAST ERROR BIAS FOR MULTIPLE-STAGE PRODUCTION-INVENTORY SYSTEMS. *Decision Sciences*, *13*(4), 570–584. http://doi.org/10.1111/j.1540-5915.1982.tb01183.x

Bowersox, D. J., Closs, D. J., & Cooper, M. B. (2002). Supply chain logistics management. *McGraw-Hill*, 498. Retrieved from http://books.google.com.co/books?id=cnYeAQAAIAAJ

Chase, C. W. (2013). Demand-Driven Forecasting. *Statewide Agricultural Land Use Baseline 2015*, *1*, 386. http://doi.org/10.1017/CBO9781107415324.004

Cleophas, C., Frank, M., & Kliewer, N. (2009). Simulation-based key performance indicators for evaluating the quality of airline demand forecasting. *Journal of Revenue and Pricing Management*, *8*(4), 330–342. http://doi.org/10.1057/rpm.2009.17

Fildes, R., & Goodwin, P. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning, (November 2016). http://doi.org/10.1016/j.ijforecast.2008.11.010

Fildes, R., & Kingsman, B. (2005). Demand Uncertainty And Lot Sizing In Manufacturing Systems: The effects of forecasting errors and mis-specification. *The Department of Management Science*, 1–33.

Fildes, R., & Kingsman, B. (2011). Incorporating demand uncertainty and forecast error in supply chain planning models. *Journal of the Operational Research Society*, *62*(3), 483–500. http://doi.org/10.1057/jors.2010.40

Freedman, D., Pisani, R., & Purves, R. (2007). *Statistics* (4th editio). W. W. Norton & Company.

Gelman, a, & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. *Policy Analysis*, 1–651. http://doi.org/10.2277/0521867061

Hausman, W. (2004). Supply Chain Performance Metrics. *The Practice of Supply Chain Management: Where Theory and Application Converge*, 61–73. http://doi.org/10.1108/01443571211195727

Hoover, J. (2009). How To Track Forecast Accuracy To Guide Forecast. *Foresight*, *summer*(14), 17–24.

Hyndman, R. (2014). Another look at estimators for intermittent demand.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. http://doi.org/10.1016/j.ijforecast.2006.03.001

Johnson, J., Tellis, G. J., & Ip, E. H. (2013). To Whom , When , and How Much to Discount ☐? A Cons Discounts. *Journal of Retailing*, *89*(4), 361–373. http://doi.org/10.1016/j.jretai.2013.08.002

Johnston, F., Taylor, S., & Oliveria, R. (1988). Setting company stock levels. *Journal of the Operational Research …*, *39*(1), 15–21. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract%5Cnhttp://www.jstor.org/stable/10.2307/2581994

Kerkkänen, A., Korpela, J., & Huiskonen, J. (2009). Demand forecasting errors in industrial context: Measurement and impacts. *International Journal of Production Economics*, *118*(1), 43–48. http://doi.org/10.1016/j.ijpe.2008.08.008

Krajewski, L. J., Ritzman, L. P., & Mallhotra, M. K. (2013). *Operations*

Management: Processes and Supply Chains. http://doi.org/10.15358/9783800644858

Lee, T., Cooper, F., & Adam, E. (1993). The effects of forecasting errors on the total cost of operations. *Omega*, *21*(5), 541–550. http://doi.org/10.1016/0305-0483(93)90022-D

Lee, T. S., & Shih, W. (1989). Optimal forecast biasing in theoretical inventory models. *International Journal of Production Research*, *27*(5), 809–830. http://doi.org/10.1080/00207548908942589

Leitch, G., & Tanner, J. E. (1991). Economic Forecast Evaluation: Profits Versus the Conventional Error Measures. *The American Economic Review*, *81*(3), 580–590. http://doi.org/10.2307/2006520

Levy, M., & Weitz, B. (2011). *Retailing Management*. McGraw-Hill Education.

Lu, R. M. (2014). *A Study to Examine the Importance of Forecast Accuracy to Supply Chain Performance , the Contributing Factors and the Improvement Enablers in Practice*.

Mendenhall, W., & Sincich, T. (2011). *A Second Course in Statistics: Regression Analysis* (7 edition). Pearson.

Ott, K., Mensendiek, A., & Gmeinwieser, K. (2013). A framework for economic demand forecast evaluation. *OR Insight*, *26*(3), 203–218. http://doi.org/10.1057/ori.2013.3

Ritzman, L. P., & King, B. E. (1993). The relative significance of forecast errors in multistage manufacturing. *Journal of Operations Management*, *11*(1), 51–65. http://doi.org/10.1016/0272-6963(93)90033-L

Sanders, N. R., & Graman, G. A. (2009). Quantifying costs of forecast errors: A case study of the warehouse environment. *Omega*, *37*(1), 116–125. http://doi.org/10.1016/j.omega.2006.10.004

Sheather, S. (2009). *A Modern Approach to Regression with R. Design* (Vol. 102). New York, NY: Springer New York. http://doi.org/10.1007/978-0-387-09608-7

Stadtler, H., & Kilger, C. (2004). *Supply chain management and advanced planning*. Retrieved from http://link.springer.com/content/pdf/10.1007/978-3-540-74512-9.pdf

Strijbosch, L. W. G., & Moors, J. J. A. (2005). The impact of unknown demand parameters on (R,S) -inventory control performance. *European Journal of Operational Research*, *162*(3), 805–815. http://doi.org/10.1016/j.ejor.2003.09.026

Syntetos, A. A., Nikolopoulos, K., & Boylan, J. E. (2010). Judging the judges through accuracy-implication metrics: The case of inventory forecasting. *International Journal of Forecasting*, *26*(1), 134–143. http://doi.org/10.1016/j.ijforecast.2009.05.016

Wemmerlöv, U. (1989). The behavior of lot-sizing procedures in the presence of forecast errors. *Journal of Operations Management*, *8*(1), 37–47. http://doi.org/10.1016/S0272-6963(89)80004-X

Xie, J., Lee, T. S., & Zhao, X. (2004). Impact of forecasting error on the performance of capacitated multi-item production systems. *Computers and Industrial Engineering*, *46*(2), 205–219. http://doi.org/10.1016/j.cie.2003.12.020

Zhao, X., Xie, J., & Wei, J. C. (2002). The Impact of Forecast Errors on Early Order Commitment in a Supply Chain. *Decision Sciences*, *33*(2), 251–280. http://doi.org/10.1111/j.1540-5915.2002.tb01644.x