Forecasting the Product Return Rate in the e-commerce context

Marta Filipa Pereira Reis

Dissertation written under the supervision of
Professor Pedro Afonso Fernandes

Dissertation submitted in partial fulfillment of requirements for the MSc in Business Analytics, at the Universidade Católica Portuguesa, 03.01.2024.

Forecasting the Product Return Rate in the e-commerce context

Marta Filipa Pereira Reis

Abstract

The main objective of this thesis is to understand and analyze the impact of product returns on the losses of e-commerce companies. To this end, data from an e-commerce company in the fashion sector was used. The variables analyzed were: product return rate, company's losses, and company's profit.

Studying the relationship between these three variables revealed, on the one hand, that the higher the product return rate, i.e., the percentage of products returned, the higher the company's losses, i.e., the costs the company incurs when its customers return the product they bought. On the other hand, the higher the company's losses, the lower the profits.

Therefore, there is a significant impact and a relationship between returned products and company losses. As a result, e-commerce companies should be concerned and carry out periodic monitoring, as well as develop strategies to reduce the percentage of returned products to reduce company losses.

One way of doing this is by using forecasting methods. With this in mind, I studied and analyzed four forecasting models: HP filter, MED filter, AR(1) model, and Theta Model.

Among all these models, the best-performing model was the Theta Model, regardless of the variable under study. Therefore, this model was applied to make the monthly forecast for next year. According to the forecast values, we concluded that if the company doesn't implement any strategy to ensure a reduction in returned products, then the company's profits will fall.

Keywords: Time series; Forecasting models; Product return rate; E-commerce; Company's losses.

Forecasting the Product Return Rate in the e-commerce context

Marta Filipa Pereira Reis

Resumo

O principal objetivo desta tese é compreender e analisar o impacto das devoluções de produtos nas perdas das empresas de comércio eletrónico. Para tal, foram utilizados dados de uma empresa de comércio eletrónico no sector da moda. As variáveis analisadas foram: taxa de devolução de produtos, prejuízos e lucros da empresa.

Quanto maior a taxa de devolução de produtos, ou seja, a percentagem de produtos devolvidos, maiores são as perdas da empresa, ou seja, os custos que a empresa tem quando os seus clientes devolvem o produto que compraram. Por outro lado, quanto maiores são as perdas da empresa, menores serão os seus lucros.

Existe um impacto significativo e uma relação entre os produtos devolvidos e as perdas da empresa. Assim sendo, as empresas de comércio eletrónico devem preocupar-se e efetuar um acompanhamento periódico, bem como desenvolver estratégias para reduzir a percentagem de produtos devolvidos, de modo a diminuir os prejuízos da empresa.

Uma das formas de garantir o controlo e perceber o irá acontecer nos próximos meses é através da utilização de métodos de previsão. Com este objetivo, estudei e analisei quatro modelos de previsão: Filtro HP, Filtro MED, Modelo AR(1) e Modelo Theta.

O Modelo Theta teve melhor desempenho, independentemente da variável em estudo. Assim, este modelo foi aplicado para efetuar a previsão mensal para o próximo ano. De acordo com os valores da previsão, concluímos que se a empresa não implementar nenhuma estratégia para garantir uma redução dos produtos devolvidos, então os lucros da empresa irão diminuir.

Palavras-chave: Séries temporais; Modelos de previsão; Taxa de devolução de produtos; Comércio eletrónico; Prejuízos da empresa.

Contents

List of Figures

List of Tables

1 Introduction

E-commerce has become a fundamental part of the economy over the years. Its ability to allow businesses to reach a broader range of customers and enable them to shop at their convenience has contributed to its growth and significance. The use of electronic networks for business increased in the 1980s and boomed in the 1990s with the emergence of online market platforms like Amazon and eBay. The practicality of e-commerce has led it to become the fastest-growing retail market. Electronic commerce is considered phenomenal today as its powerful concepts and processes have profoundly changed our lives (Taher, 2021).

The COVID-19 pandemic has profoundly impacted our lives, affecting health, the economy, and shopping habits, leading to a surge in e-commerce. Between 2020 and 2021, visiting a physical store was restricted, leading to a significant increase in digital activities. Customers have turned to the online market to make purchases, increasing the number of returns. The consumer journey has been changing, resulting in a widespread increase in the use of e-commerce (Higueras-Castillo et al., 2023).

In an ideal world, companies would not have to deal with product returns. However, returns are a common struggle that is sometimes underestimated. The financial impact of product returns is only part of the issue. There are also many environmental consequences, such as increased greenhouse gas emissions, packaging waste, unnecessary energy use, and increased transportation. Vogue Business reports that

> "US returns alone generate 5 billion pounds of landfill waste and 15 million tonnes of carbon emissions annually. This is equivalent to the amount of trash produced by 5 million people in a year." (SCHIFFER, 2019)

Improving the consumer experience while understanding the most effective strategies to achieve an optimum financial return is one of the biggest challenges for many companies. With the growing e-commerce sector, the economic and environmental ramifications associated with product returns are even more complex.

By understanding the financial impact of returns, companies can optimize their stocks, improve customer service, and reduce logistics costs. Moreover, analyzing trends and forecasting returns can help companies make informed decisions about product quality, marketing strategies, and operational efficiency (Xiaofeng and Tijun, 2009). This proactive approach contributes to the company's financial health, enhances the customer experience, and promotes sustainable business practices.

Therefore, my research focuses on investigating the relationship between the amount of product returns and the company's losses in the e-commerce fashion context. Additionally, I aim to analyze and implement some forecasting models and methods to predict the likelihood of product return rate, company losses, and profits. To achieve this goal, I will address the following research questions: Does a rise in product return rate lead to increased losses in e-commerce? Which forecasting model is the most effective in predicting the probability of product returns and company profits and losses?

To provide answers to this question, I conducted my research in collaboration with an e-commerce fashion company and analyzed their data. The company offers online shopping services and has gained international recognition for its user-friendly website and quality handmade products. The company aims to empower women to feel good and confident about their looks through clothing and beauty products. They also prioritize sustainable fashion and reducing their ecological footprint. All of their items are handmade to ensure high quality and durability. Customer satisfaction is the company's top priority, and it strives to maintain direct contact with its customers through various digital platforms.

Despite having a powerful marketing strategy, the company is continuously growing and trying to find new ways to provide its customers with an exceptional experience. This research aims to analyze the impact of product returns on company losses and identify the best forecasting method or model to minimize losses.

I found exploring the HP filter, MED filter, AR(1) model, and Theta model pertinent to this research. All these models arouse much curiosity from researchers, and there are many opinions on which of them can get more accurate forecasts.

The research includes a literature review that explains and describes the relevance and impact of product returns on company losses. In addition, I explored the importance of consumer satisfaction and, consequently, consumer behavior to understand what leads a consumer to return their product. Then, I strive to describe and analyze the data from the e-commerce company to have an overview before deciding on each model to apply. I introduce the four models in the forecasting models chapter. A methodology is proposed to explore the forecast workflow and understand how to choose the best model to make the predictions for the following year. Once we have applied the best model, I present the results obtained in the Findings chapter. In the discussion, some constraints and limitations are noticed. Finally, the last chapter presents the conclusions of this dissertation, as well as some suggestions for further research in this area of study.

2 Literature Review

This chapter presents a theoretical framework that tackles the main research questions and study purpose. The main objective is to examine how product returns impact company losses in e-commerce and understand the reasons behind returning a product. I have reviewed relevant literature, including academic journals, books, and news articles.

The initial part of this literature review begins by setting the stage with an overview of the e-commerce landscape. Then, I delve into the critical relationship between product returns and company financial losses, emphasizing their significant business implications. Additionally, I explore the influence of consumer behavior and satisfaction on return rates.

I present a comprehensive conceptual model that illustrates the connection between e-commerce product returns, company losses, and consumer behavior. Finally, I will explore the significant contribution of forecasting methods toward reducing product returns.

2.1 E-commerce context

According to Taher (2021), e-commerce involves using telecommunication networks to automate business workflow and relations. The percentage of e-commerce in total retail sales worldwide is set to continue growing, as seen in Figure 1. (Statista, 2021)

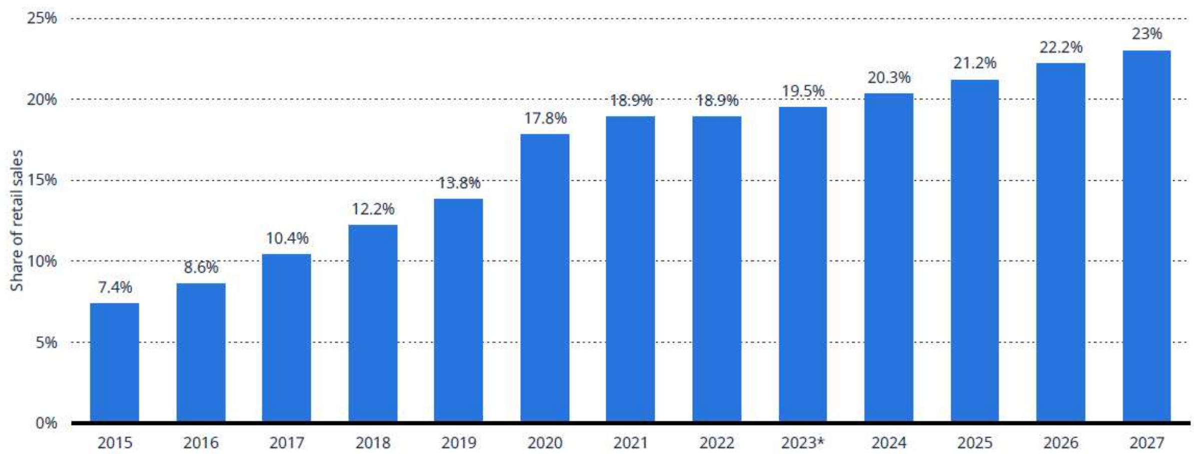


Figure 1: E-commerce as a percentage of total retail sales worldwide 2015-2027

E-commerce has, therefore, revolutionized the business landscape and is now a necessity. One of the most notable e-commerce benefits is its exceptional convenience. With a simple tap on their mobile devices, customers can easily acquire products from anywhere in the world. The shopping process is swift and straightforward, taking at most 15 minutes to complete. Within a week, the selected products are delivered to their doorsteps Taher (2021). This expeditious and user-friendly shopping experience makes e-commerce highly efficient for obtaining desired goods.

With changing customer expectations, companies must keep up by offering features and strategies to enhance the e-commerce experience. One significant issue affecting companies' success is the number of product returns.

E-commerce companies experience a substantially higher return rate of approximately 30% compared to physical retail stores, with a return rate of around 8% (Yan et al., 2022). Therefore, it is essential to analyze the impact of product return policies on e-commerce. While the number of returns a company receives may be underestimated, it can significantly impact its overall success.

2.2 Product Returns: A Business Imperative

Product returns commonly occur when a customer wishes to exchange a purchased item for store credit or a refund. In recent years, there has been a significant increase in product returns due to the rise of e-commerce businesses and the industry's overall growth. Additionally, the competitive market and faster product life cycles have also contributed to the higher flow of returned products (Agrawal and Singh, 2019).

Most companies have some form of return policy for their customers, and it is not a new practice. However, it is essential to reconsider the value of offering easy and flexible return policies. Corporate and consumer perspectives must be understood (Paswan, 2018).

From the company's perspective, reducing the costs of returning products and influencing consumer purchases through restocking fees and pricing policies is crucial. On the other hand, from the customer's perspective, it is essential to understand consumer behavior, i.e., the intentions behind returning a product (Shulman et al., 2009).

It's crucial to understand the significance of returns management. According to McKinsey research, it's concerning that a third of retailers don't consider it a top priority, and around 25% lack an effective system for managing returns. (Ader et al., 2021)

Understandably, shipping and logistics costs are a significant concern for retailers, but it's essential to recognize the impact that poor returns management can have on customer satisfaction and loyalty. As such, it's vital to approach this topic with a sense of responsibility and sensitivity (Ader et al., 2021).

2.2.1 Unraveling market trends: An in-depth analysis of product returns

To understand the impact of product returns, it's essential to keep an eye on market trends. In recent years, there has been a noticeable increase in product returns, which can be attributed to a combination of factors, including the convenience of online shopping, lenient return policies, and consumers valuing convenience over quality.

The average return rate for clothing marketplaces in the fashion industry can exceed 50 percent (Yan et al., 2022). This is concerning, as the fashion industry is one of the most significant contributors to environmental pollution due to garment disposal (McFall-Johnsen, 2020).

Amazon is a good example of this trend, with return rates ranging from 5 percent to 15 percent, depending on the product category. In some categories, like electronics and clothing, the return rate can be as high as 40 percent (Williams, 2021). According to Statista, clothing is the category with the highest percentage of returned products in Europe (Chevalier, 2023).

It's common for customers to take advantage of the offer of free returns when shopping online, even if they haven't physically seen or touched them. This facility plays a vital role in the increase of product returns. This is especially true when it comes to clothing. People are turning their homes into fitting rooms; they want to try things on in the comfort of their homes and return anything that doesn't meet their expectations (Calma, 2019).

Shopping online is convenient, but has led to increased returns and impulsive buying. However, it has also made it easier to experiment with new trends and products (Paswan, 2018). While this policy can boost consumer satisfaction, it presents challenges for companies by trying to identify whether it's a problem or a competitive advantage.

2.2.2 Reverse Logistics as a Way to Understand Product Return Trends

Reverse logistics involves collecting, disassembling, and processing used products, parts, and materials to ensure a sustainable recovery (Shulman et al., 2009). It is also known as the management of product returns from consumers back to manufacturers, sellers, and suppliers (Kammerer et al., 2020).

It is crucial in optimizing resources and reducing costs associated with increased returns. However, the current economic scenario has resulted in higher prices for returns, which can affect the company's profitability (Ramírez, 2012).

Reverse logistics plays a crucial role in optimizing resources and reducing costs associated with increased returns. To minimize these costs, firms make decisions that maximize the value of salvaged products.

According to Statista, the global reverse logistics market is projected to continue experiencing growth, as we can see in Figure 2:

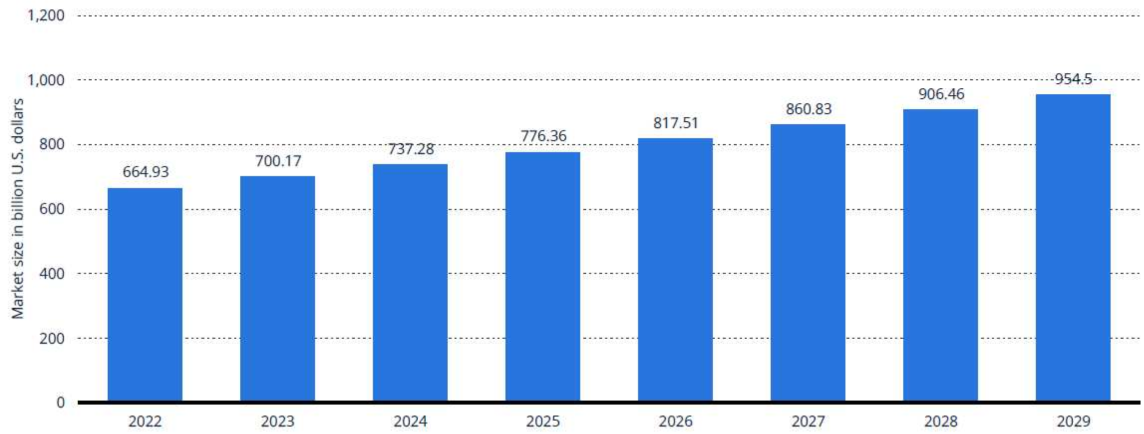


Figure 2: Global reverse logistics market size forecast (billion dollars)

By having a deeper understanding of what is reverse logistics, we can gain a deeper understanding of market trends related to product returns. As depicted in Figure 2, it is evident that the forecasted size of the global reverse logistics market is continuously expanding, aligning with the upward trajectory of product returns.

This highlights the crucial importance of understanding how product returns affect the health of e-commerce businesses.

2.3 Company losses

The impact of return policies goes beyond just satisfying customers. It also has a significant impact on a company's ability to adapt to market needs, operational efficiency, and brand reputation. In addition, we can't ignore the environmental impact of returned products (Shulman et al., 2009).

Optoro, a return logistics company, reports that over a quarter of returned items end up being discarded. This is exemplified by Amazon's practice of distributing millions of returned items still in excellent condition (Zhang et al., 2023).

Each returned package leaves behind a trail of emissions from transportation, regardless of the carrier used, contributing to air pollution and climate change. Sadly, most of these discarded items end up in landfills, with about 5 billion pounds of returned products ending up as trash in the US (Calma, 2019).

According to Lee and Morewedge (2023) research, consumers returned 16.5% of their total purchases in 2022, causing retailers to lose an estimated 816 billion dollars. This return rate is similar to 2021, when consumers returned 16.6% of their purchases, resulting in an estimated loss of approximately 761 billion dollars for retailers. It is essential for retailers to find ways to reduce return rates and minimize revenue loss.

A major downside is that returned products can often no longer be sold at their original price. Deloitte's study reveals that items kept by customers for a longer period of time lose their value, leading companies to offer discounts. Figure 3 depicts the correlation between time and value, where we can see how the discount amount varies based on time elapsed. (Kammerer et al., 2020)

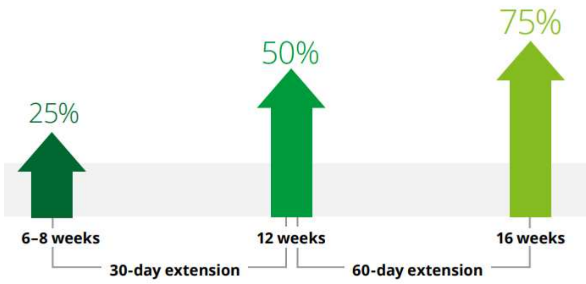


Figure 3: The traditional markdown cadence for fashion

As e-commerce continues to grow and free returns become the norm, the environmental impact will only worsen. McKinsey reports that 58% of retailers do not have a dedicated person responsible for minimizing returns and managing their impact. Having someone focused solely on this aspect could be a worthwhile strategy for companies.

According to Petersen and Kumar (2009) return policies can have a significant impact on how customers decide whether or not to return a purchased product. When retailers establish their return policies, they need to balance the customer experience, competitive landscape, and profit margins to determine the appropriate time frame for returns (Kammerer et al., 2020).

Petersen and Kumar argue that implementing a restrictive return policy may have negative consequences, such as reducing customer satisfaction and negatively influencing customer decision-making (Petersen and Kumar, 2009). On the other hand, having a flexible return policy may also lead to abuse and impulsive behavior from consumers, which may prove to be disadvantageous (Bechwati and Siegal, 2005).

Based on research conducted by Petersen and Kumar in 2009, if there is a negative correlation between a customer's future value and their product return behavior, managers may attempt to discourage returns in order to maximize profits for the company (Petersen and Kumar, 2009). However, Reinartz and Kumar's study from 2003 suggests that this relationship can be positive to some degree. This highlights the challenge that managers face in effectively managing returned products (Reinartz and Kumar, 2003).

Did you know that according to a study by McKinsey, the vast majority of people find it crucial to have a free return policy? This is because it helps to strengthen the relationship between the company and the consumer, and ensures customer satisfaction. It's interesting to note that while most people consider the return of products to be a necessary evil, over time, loyal consumers tend to adapt to the return process, especially when it comes to clothing. However, companies may start charging for return transportation, which can have a negative impact on their customers (Ader et al., 2021).

In addition, Maxham and Bower have concluded that customers who are given the option of a free product return are more likely to make future purchases than those who have to pay a fee for returns (Petersen and Kumar, 2009).

There is also a study that focused on whether a return policy can benefit a company financially. The study involved a large sample of 26,000 e-commerce clients over a period of six months, and the findings were quite intriguing. The study concluded that a company that is aware of the costs associated with product returns can significantly reduce the total amount of money spent on such returns. In fact, the study revealed a nearly 7% decrease in product returns. Additionally, they found that companies that allocate resources to minimize the costs of product returns are able to protect their profits. Finally, the study found that when companies consider both consumer behavior and the costs of product returns, the number of product returns can decrease by as much as 25% (Petersen and Kumar, 2015).

To summarize, a study conducted by Deloitte identified the top five issues that lead to company losses due to product returns. This negative consequence is illustrated in Figure 4 (Kammerer et al., 2020).

Figure 4: Five issues that lead to company losses - Deloitte study

To conclude, product returns have a huge impact on company losses, so it's important for companies to strike a balance between cost allocation and consumer satisfaction to maintain a positive reputation and retain loyal customers (Petersen and Kumar, 2010). As a result, we will analyze the importance of consumer satisfaction for e-commerce and how it affects consumer behavior.

2.4 Consumer Satisfaction

In strategic marketing, consumer satisfaction is a crucial concept. It refers to the emotional response that consumers feel when evaluating a product they have consumed (Hadiantini et al., 2021). Understanding the relationship between product return experience, return satisfaction, and customer outcomes can help companies create more informed returns management strategies.

Satisfied customers tend to exhibit loyal behavior, favoring the firm over its competitors, being less price-sensitive, and attracting new customers through word-of-mouth. Ultimately, customer satisfaction through loyalty leads to higher revenues and better financial performance (Eklof et al., 2020).

Research conducted by Hadiantini et al. (2021) shows that service quality, brand image, customer value, user experience, customer experience, trust, promotion, and price are the most significant factors influencing customer satisfaction, indicating that e-commerce consumers prioritize the services provided.

In addition, the ACSI model was discussed as a measure to test firms' performance, as seen in Figure 5.

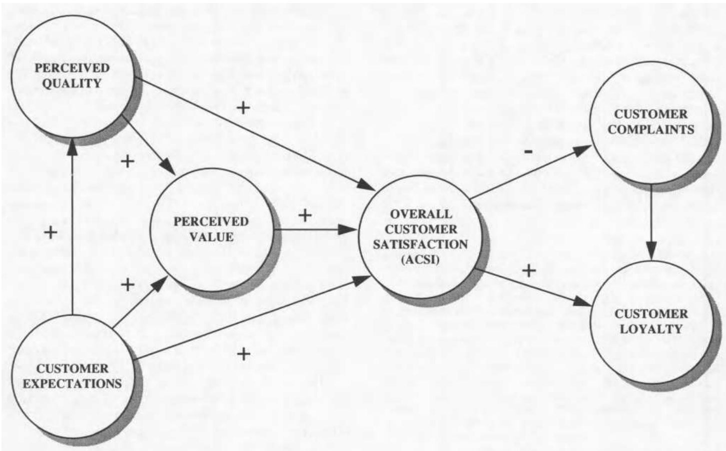


Figure 5: ACSI Model

The ACSI model has three antecedents that characterize it: perceived quality, perceived value, and customer satisfaction. To ensure customer satisfaction, product quality, which is the assessment and feedback of the product by consumers, is expected to impact customer satisfaction positively. Perceived value is the relationship between quality and price and is generally associated with higher customer satisfaction. The market's expectations, which relate to the company's offer and word-of-mouth marketing, are also crucial in predicting customer satisfaction and building brand-loyal customers (Fornell et al., 1996).

Furthermore, customer expectations should be positively related to perceived quality and, therefore, perceived value. On the other side of the figure, we have the importance of customer complaints and loyalty. According to Fornell et al. (1996), if customers are satisfied, there will be fewer complaints and more customer loyalty.

Schiffman et al. (2014) concluded that a one-point increase in customer satisfaction could result in an 11.5 percent increase in net income for companies.

2.5 Consumer Behaviour

It is critical to comprehend the factors that influence customers to return products and the circumstances under which they do so. This understanding will assist in decreasing return rates and enhancing customer satisfaction. Every time a customer returns a product, it presents an opportunity to gain insight into their needs and preferences, which can enable businesses to tailor their offerings to better meet those needs and drive future sales (Kammerer et al., 2020).

According to Petersen and Kumar (2009), to operate successfully, it's important to understand that consumer behavior and preferences are constantly evolving, and these changes can significantly impact marketing strategies.

By having a deep understanding of what motivates consumers to buy, what needs they are trying to fulfill, and what influences their product and brand choices, companies can create more effective marketing campaigns. It's now easier than ever for marketers to track and analyze consumer behavior due to technological advancements. However, it's vital to consider the reasons behind product returns (Schiffman et al., 2014).

Paswan and Pei suggested that there are two different types of consumer behavior: legitimate and opportunistic return behavior. Legitimate return behavior is typically driven by factors like a desire for uniqueness, perceived risk and consumer impulsiveness. On the other hand, opportunistic returns occur when there is immorality and self-monitoring involved (Paswan, 2018).

Therefore, we will strive to gain a better comprehension of the factors influencing consumer behavior in regard to returning products, taking into account these two distinct types of behavior (Thomas, 2023).

Misinformation is a major reason why consumers feel the need to return a product. Many times, consumers end up buying something completely different from what they thought they were purchasing due to superficial product descriptions. This can lead to a loss of trust in the brand. Visual content plays a crucial role for consumers when making purchasing decisions (Calma, 2019).

To avoid this we have a practical example of Shein, a leading e-commerce application in the United States, surpassing Amazon, although it has been facing significant controversy and ethical concerns (Neiva, 2023). Shein has implemented a clever business strategy by showcasing photos of people wearing their products. This gives potential customers confidence and ultimately boosts sales. Additionally, by showcasing a diverse range of individuals with varying body types, the number of returns is reduced.

Sundar and Noseworthy (2014) concluded that consumers only take about five to seven seconds to evaluate a product on the shelf and make a purchasing decision within three to seven seconds. This means that packaging has become an incredibly important marketing tool. According to Silayoi and Speece (2007), packaging design is used to catch the consumer's eye and create expectations for the product in their mind when they are not familiar with it. It's fascinating how much thought goes into packaging design to influence our purchasing behavior.

Another common issue, especially with online shopping, is when the size of the product doesn't match the consumer's expectations. This often leads to the need for exchanges.

Zara offers two helpful features when purchasing a product. Firstly, they have a size guide to help buyers choose the right size. Secondly, they give customers the option to answer a few questions about their height, weight, and preferred fit (tight, perfect, or loose) to find the best size for them. Zara also provides statistics on the percentage of people with similar characteristics who purchased the product and did not return it. This ensures a well-informed consumer experience, and customer satisfaction, and reduces the number of returns.

When consumers pay for a service, they expect to receive a product that is not damaged or faulty. When this doesn't happen, it can really harm the brand-consumer relationship. Companies can avoid this situation by checking the packaging process to ensure that products arrive safely.

Late delivery is also a common reason for poor customer satisfaction. Consumers rely on companies to deliver products on time, and when this doesn't happen, it can lead to frustration and lost trust. To avoid this issue, companies should provide specific information about shipments, indicate deadlines, notify customers if orders are delayed, and do their best to meet deadlines (Thomas, 2023).

Another issue that frustrates consumers is receiving the wrong product from a company. It's important for companies to have a fair and customer-friendly return policy in order to ensure customer satisfaction. In fact, it's considered unethical to charge customers for items they didn't purchase. So, if you do receive the wrong product, the company should reimburse you and you're not obligated to return the item (Thomas, 2023).

A common reason for returns is during the holiday season when customers receive gifts that they don't want or can't identify with. In fact, in 2019, over 41.6 billion euros worth of items were returned during the months of November and December alone (Thomas, 2019). Some companies have adjusted their return policies during this time to offer gift cards instead of returns in order to attract customers and ensure their satisfaction.

It's understandable that consumers may feel the need to return their purchases for legitimate reasons. However, there are some cases where people engage in opportunistic behavior that includes fraud. Unfortunately, return fraud is a costly issue for the retail industry.

According to the National Retail Federation, it's estimated that return fraud costs the industry around 13 billion dollars in lost sales (NRF, 2021). Shockingly, around 40% of buyers admit to using items with the intention of returning them, which can include hiding the tags, using the product, and then returning it when they no longer need it (DUHS, 2020). This behavior can be very harmful to companies, as less than half of returned products can be resold at their original price (Rossi, 2022).

In conclusion, considering all these factors that encourage a consumer to return a product, companies are becoming more aware of which strategies to implement.

2.6 The importance of forecasting

Forecasting could be relevant in business context, allowing us to make operational and strategic decisions, such as planning financial requirements, capacity and production planning, and identifying market opportunities.

Forecasting product returns plays a fundamental role in the reverse logistics network; it helps companies gain some insight into consumer behavior and plan new measures to reduce the number of returns (Xiaofeng and Tijun, 2009).

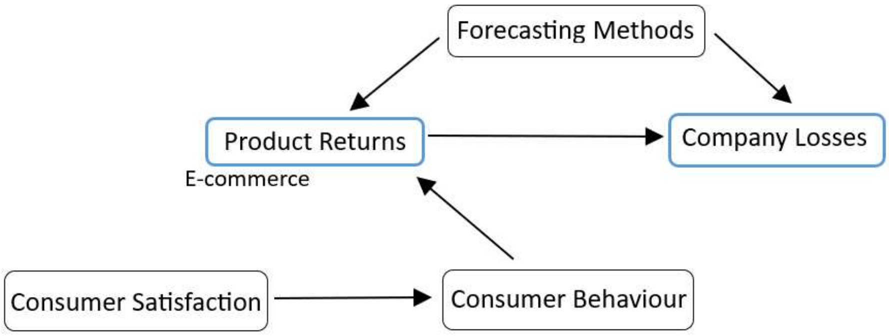
To make a forecast, many different methods and models are available, so it is essential to analyze which model or method may provide a more sophisticated outperform for our specific data and context (Doering and Suresh, 2016).

Before applying any forecasting method, it's important to conduct pre-processing of the data. This includes checking for missing values and ensuring accuracy. Some models may also require more detailed attention regarding seasonality, so keeping that in mind is essential. Managing returns in e-commerce can be challenging, and doing it profitably is particularly difficult. Therefore, minimizing losses is crucial, regardless of the business size. To tackle this problem, companies can implement proactive measures such as providing detailed product descriptions and comprehensive sizing charts, improving product quality, and utilizing AI-powered tools to predict the likelihood of products being returned.

According to Zikopoulos (2017), the quality and quantity of returned products can significantly affect reverse logistic planning and performance. One effective way of reducing uncertainty in this regard is to forecast product returns, which can help establish an efficient reverse logistic system (Agrawal and Singh, 2019).

To conclude, companies can minimize losses and create a positive customer experience by carefully analyzing past data, understanding the relationship between product returns and company losses, and forecasting the likelihood of returns. This could help them make informed decisions and take proactive measures to prevent returns, ultimately enhancing their bottom line (Petropoulos et al., 2022).

2.7 Full Conceptual Model



In this conceptual model, we can quickly see all the topics addressed. The main variables under study are product returns and company losses, and the main objective is to understand the impact of product returns on a company's financial health in the context of e-commerce. To this end, I will use forecast models to understand the likelihood of product returns and company losses the next year; hence, I have two arrows pointing to each variable. On the other hand, I found it interesting to address the theme of consumer satisfaction, as this will lead to a particular action, i.e., consumer behavior, that can be a significant influencer of the increase of product returns and consequent company losses.

3 Data description

In this chapter, I will explore the dataset a fashion e-commerce company provided for my thesis. It contains 94 observations and three different variables from January 2016 to October 2023 - return_rate, profit and cost_return.

To carry out a more exploratory analysis of the data, I decided to use the R programming language and the Gretl application. For the best performance of the forecast models, it is recommended to eliminate seasonality from the series. As a result, I used the Gretl application, see https://gretl.sourceforge.net/ for details, and implemented the TRAMO-SEATS option with the output seasonally adjusted series. As an output, it gave me a CSV file with my three adjusted variables - return_r_sa, profit_sa and cost_ret_sa.

In Table 1, we can see a statistical overview of the initial and adjusted variables; this way, we better understand each variable in the study.

Statistic	N	Mean	St. Dev.	Min	Max
return_rate	94	7.525	2.356	4.050	12.500
cost_return	94	2.339	1.406	0.707	5.393
profit	94	37.070	10.424	10.400	55.823
return_r_sa	94	7.491	2.035	4.935	10.699
cost_ret_sa	94	2.298	1.262	0.786	5.173
profit_sa	94	37.273	9.453	11.389	53.254

Table 1: Summary Statistics

In Figure 6, we have the histogram of the adjusted variables, a widespread statistical analysis, as it allows us to explore the distribution of each variable. By interpreting the figure, we can easily see all variables have an asymmetrical distribution and the distribution of the company's profit is a negatively skewed distribution.

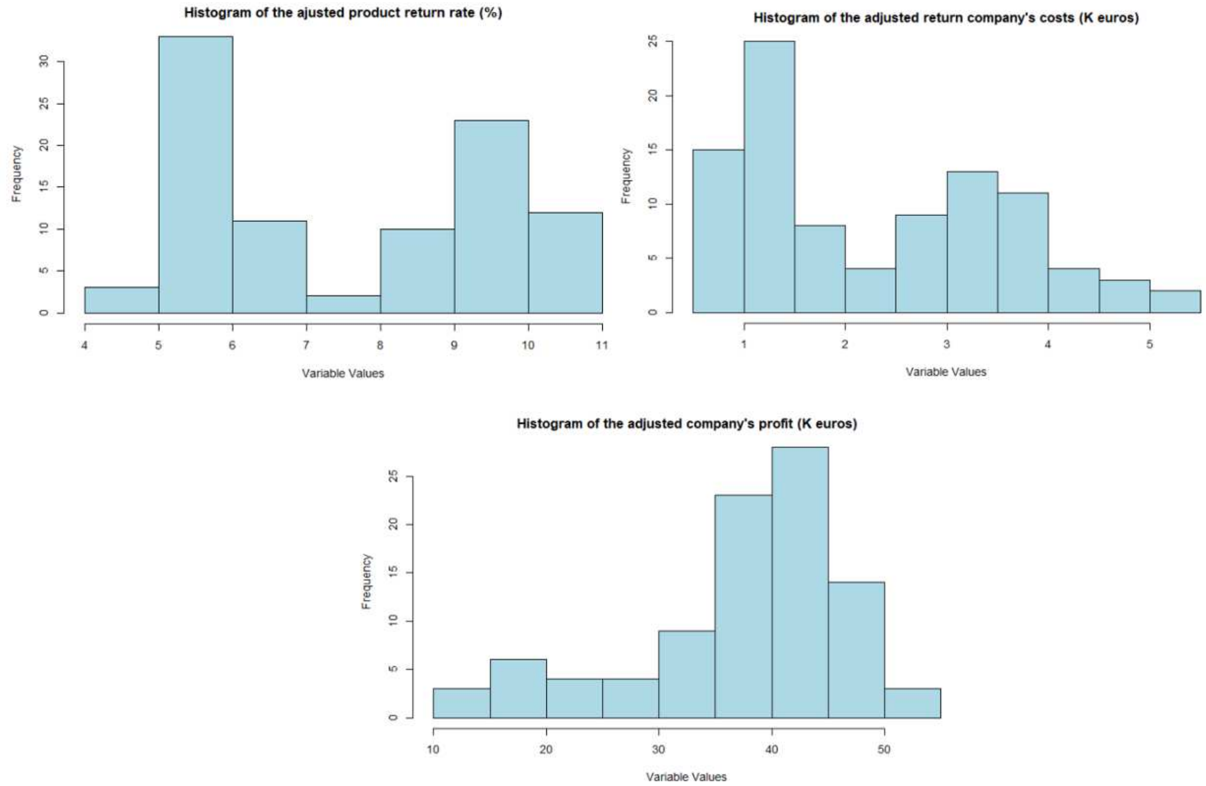


Figure 6: Histogram analysis

3.1 Product return rate - return_rate & return_r_sa

The product return rate is a monthly metric that indicates the percentage of products returned compared to those purchased. The average product return rate for the given period is 7.5%. The maximum percentage of return that the company has was 12.5%, and the minimum was 4.1%.

In Figure 7, we can compare the initial return rate variable and the adjusted return variable. We can see an increasing trend, meaning customers are returning more over time. The increase happened near 2020, which can be justified by COVID-19; people started buying more online due to the impossibility of buying in person, so the number of returns has also increased. The month that appears to have a higher rate is January. These values are often associated with Christmas when people exchange gifts, but sometimes those gifts are unsatisfactory, so January is a time for returns.

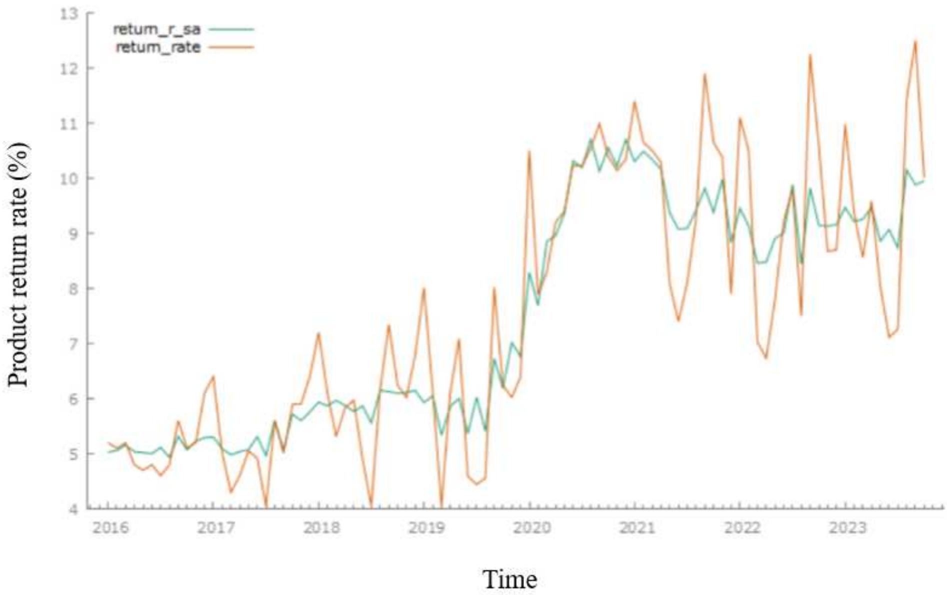


Figure 7: Adjusted product return rate vs product return rate (%)

3.2 Cost return - cost_return & cost_ret_sa

The cost of returning a product is related to the company losses in thousands of euros (K euros); check section 2.3. It includes all the costs associated with the return process, such as transportation, processing, storage, and product value loss. In the case of a product return, the company covers the cost of delivering and picking up the product from the customer's home at no extra cost. The average cost of returning per month is 2.3 thousand euros, but it increased over time, following the pattern of the return rate.

In Figure 8, we can compare the initial and adjusted cost return variables.

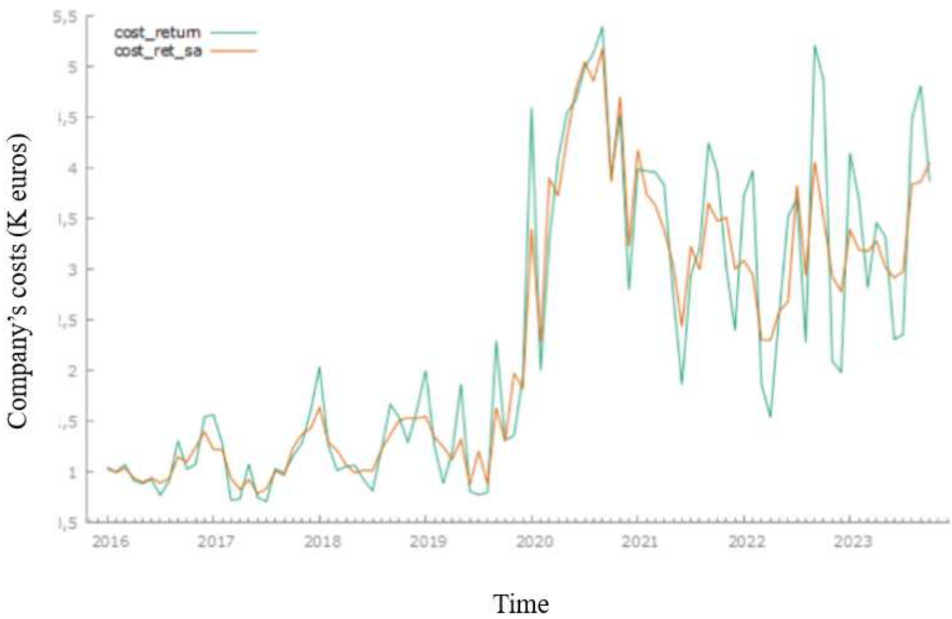


Figure 8: Adjusted cost return vs cost return (K euros)

The behavior of our time series is very similar when comparing Figure 7 and 8; each might indicate that the return rate and the cost return rate are positively correlated.

3.3 Profit - profit & profit_sa

The profit variable provides the monthly profit in thousands of euros (K euros). The average monthly profits are 37K euros.

Figure 9 compares the company's initial and adjusted profit variables. We can see a negative trend in profit over time for the company, which is not beneficial. The behavior of this series suggests a negative correlation between the product return rate and the company's profits.

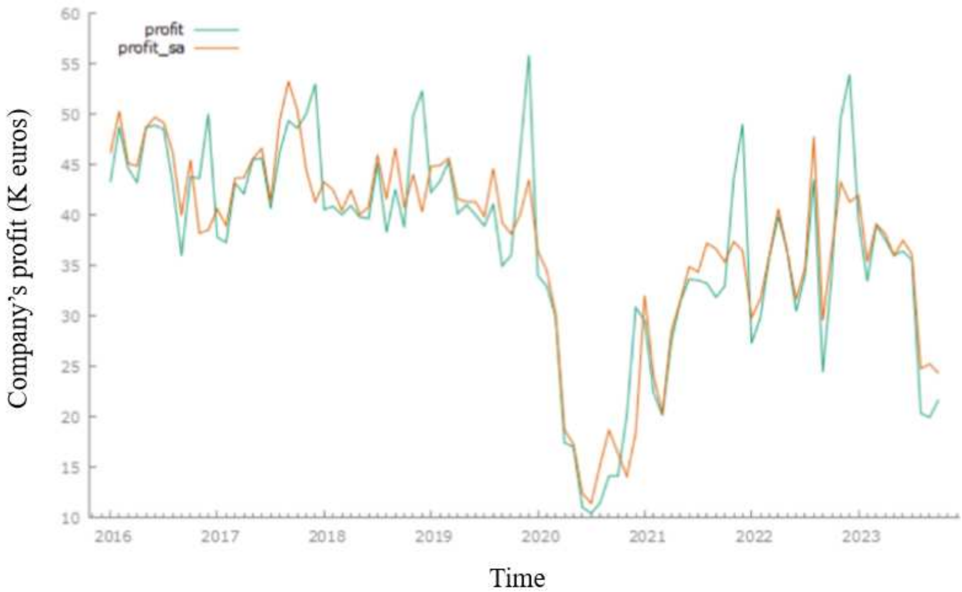


Figure 9: Adjusted profit vs profit (K euros)

3.4 Patterns analysis

It is essential to do some autocorrelation analysis when doing exploratory data analysis, especially when dealing with time series. Although we can already check some patterns in the plots above, analyzing the **ACF plots** gives us a more sensible and intuitive perception of our data patterns, which will help us choose suitable models to implement the forecast.

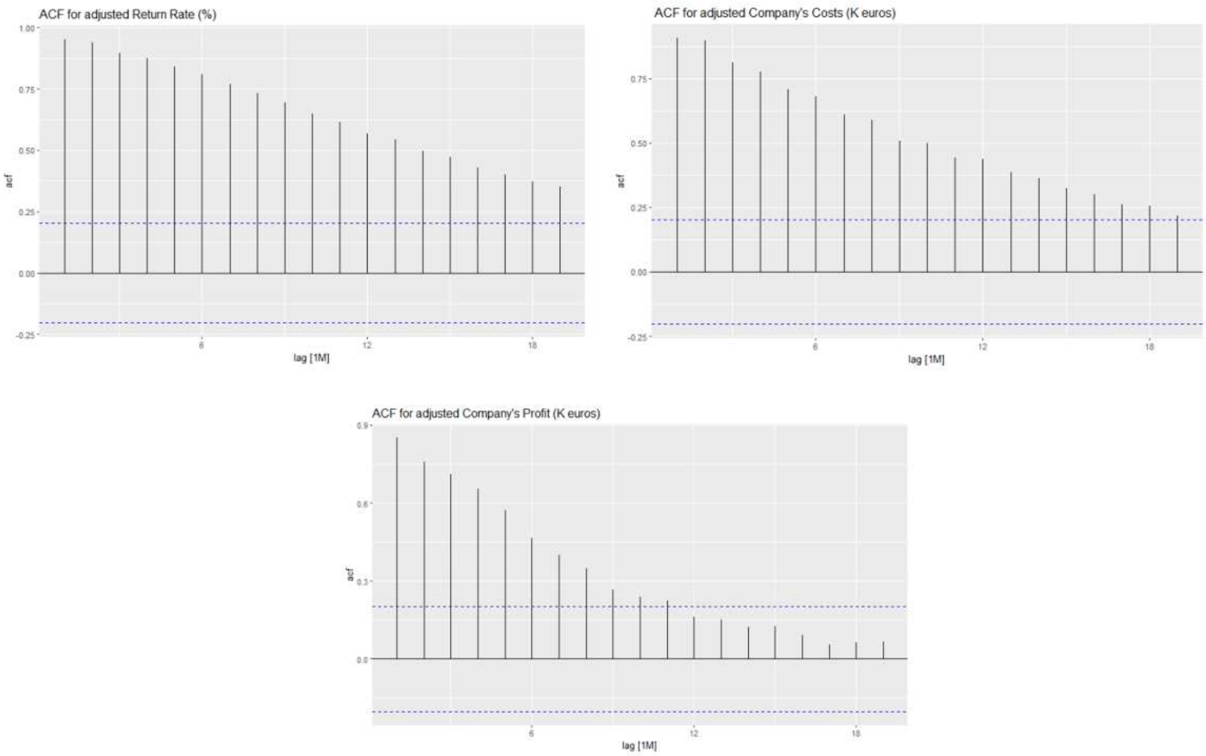


Figure 10: ACF analysis

By interpreting Figure 10, we can see that our data is not stationary since the variance and mean are not constant over time. We see a clear positive autocorrelation and a decrease in autocorrelation over time.

3.5 Relationship between adjusted variables

We can already infer the relationship between these variables from the previous plots. In Figure 11, we have three different scatter plots with a smooth method applied to study the relationship between the three adjusted variables.

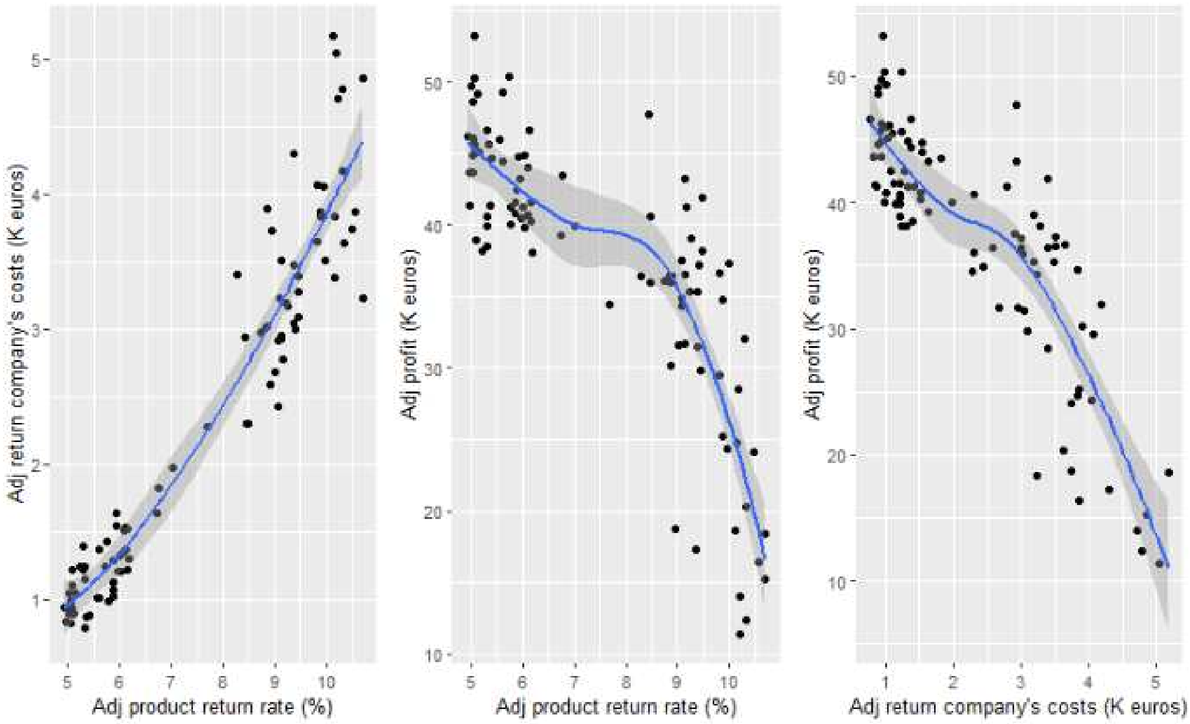


Figure 11: Relationship between adjusted variables

In addition, Table 2 shows a correlation matrix, a more quantitive way to understand the relationships.

Correlation matrix	return_r_sa	cost_ret_sa	profit_sa
return_r_sa	1	0.9533	-0.7858
cost_ret_sa	-	1	-0.8406
profit_sa	-	-	1

Table 2: Correlation Matrix

By interpreting both Figure 11 and Table 2, we can conclude that the higher the company's costs in returning the purchased products, the lower the profits. Thus, higher profits correspond to lower costs and rates of return.

4 Forecasting Models

Forecasting models play a crucial role in analyzing data and making informed decisions. In this chapter, I'll cover various forecasting models I'll apply to my data set. In the chapter dedicated to data description, chapter 3, we take a closer look at our main variables and already have some insights related to the distribution and patterns of our time series. This will help us identify which models could effectively forecast our specific data.

The four methods and models that I will study in more detail are the following: HP filter, MED filter, AR(1) model and the Theta model. In the table below, I have some reasons that motivated me to study them:

HP Filter	- Extract long term-trends - Separates time series into trend and cyclical components - Commonly used in macroeconomic analysis
MED Filter	- Attenuates noise and swift changes in time series
AR(1) Model	- Simple and easy-to-interpret model - Effective in modeling temporal relationships
Theta Model	- Simple and easy to use - Identifies more complex patterns that other simpler models are unable to identify

Forecasting practice is usually supported by time series (y_t) decomposition in trend or growth (x_t), seasonal (s_t), cyclical (c_t) and irregular (ϵ_t) components (Enders et al., 1995), that is,

$$y_t = x_t + s_t + c_t + \epsilon_t, t = 1, \ldots, n. \tag{1}$$

Usually, time series can be characterized by three different patterns, i.e., trend, seasonality, and cycle.

In fact, an one-period look ahead estimate for y_{t+1} could be obtained by doing simply

$$y^*_{t+1} = y_t + g_t + s_{t+1} - s_t, \tag{2}$$

where $g_t = x_t - x_{t-1}$ is the first difference of the trend or growth rate (%), given the data in logarithms and multiplied by 100, which is a common practice in econometrics cf. Hamilton (2018).

Thus, trend estimation becomes the main task facing the time series forecaster, apart from extracting the seasonal or intra-annual regularities of the data.

- The trend is generally seen as a slow movement over time that could be extracted with appropriate filtering techniques.

- The cycle is when the data fluctuates without a fixed frequency.
- The seasonality represents the repeated fluctuation of a specific and known period. It can be extracted with the TRAMO-SEATS, as I explained in chapter 3.

4.1 HP filter

Forecasting is typically made with trend-cycle decompositions that employ moving averages (MA) or the well-known Hodrick Prescott (HP) filter (Hodrick and Prescott, 1997). In most cases, filtering consists of applying a smooth operator M to the amoving-centered window of the data with length $l = 2k + 1$ to estimate the trend component, that is,

$$x_t = \mathrm{M}\left[y_{t-k}, \ldots, y_t, \ldots, y_{t+k}\right], t = 1, \ldots, n. \tag{3}$$

Linear and nonlinear techniques can be mobilized to perform this task. Moving Average (MA), that is, using the arithmetic mean as a smooth operator in equation (3) is a simple way to do it:

$$\tilde{x}_t = \frac{1}{2k+1} \sum_{i=-k}^{k} y_{t+i}, t = 1, \ldots, n. \tag{4}$$

A special case of moving average and linear filtering is the widely used Hodrick and Prescott (1997) filter (HP). In this method, the trend sequence $\{x_t\}$ is chosen to minimize either the sum of the squares of deviations from data series y_t or the sum of the squares of the trend's second difference, a possible measure of its smoothness:

$$\min \left[\sum_{t=1}^{N} (y_t - x_t)^2 + \lambda \sum_{t=3}^{N} (g_t - g_{t-1})^2 \right], \tag{5}$$

where $\lambda > 0$ is a penalty for the square of the difference of the trend growth $g_t \equiv x_t - x_{t-1}$, i.e. for the acceleration of the trend component x_t. The value of λ depends upon the frequency of the data e.g. $\lambda = 14400$ for monthly data. The benefit of the HP filter is that it can extract the same stochastic trend from a set of variables (Enders et al., 1995), which could be important in empirical applications.

Besides, linear filters produce smooth trends that might not capture fundamental sharp changes in the growth component of the time series under study. This problem is particularly acute with the HP method because it produces a trend with a very regular, predictable pattern and passes everything else to the cyclical component, including high-frequencies.

In a recent critique, Hamilton (2018) stressed that the HP filter produces extremely predictable cyclical components whose rich dynamics are purely artifacts created by the filter rather than reflecting any true dynamics of the data-generating process itself. Additionally, the HP filter can be extremely influenced by end-of-sample data (St-Amant and van Norden, 1997), namely, in the context of structural changes like the ones that occurred during the COVID-19 pandemic.

4.2 MED filter

As suggested by Wen and Zeng (1999), nonlinear filters may perform better than linear filters in capturing occasional, discrete shifts in the growth dynamics of economic series.

The MED filter provides a simple noise attenuation with robustness against impulsive-type movements in the data and it has proved to be effective in signaling recessions. In practice, the MED filter adopted the median X as smooth operator M in equation (3):

$$\hat{x}_t = \mathrm{X}\left[y_{t-k}, \ldots, y_t, \ldots, y_{t+k}\right] \tag{6}$$

for $t = 1, \ldots, N$ and $k = 36 - 1 = 35$ with monthly data.

The idea behind this filter is quite simple. It simply replaces each point in the time series with the median of the values in a sliding window around it. To be able to filter also the outermost observations, where the filter window partially falls outside the input signal, Wen and Zeng (1999) replicated the y_1 and y_N values as many times as needed, the so-called "first and last values carry-on appending strategy".

In Figure 12, we can have a better notion of how the median filter works. (Stone, 1995)

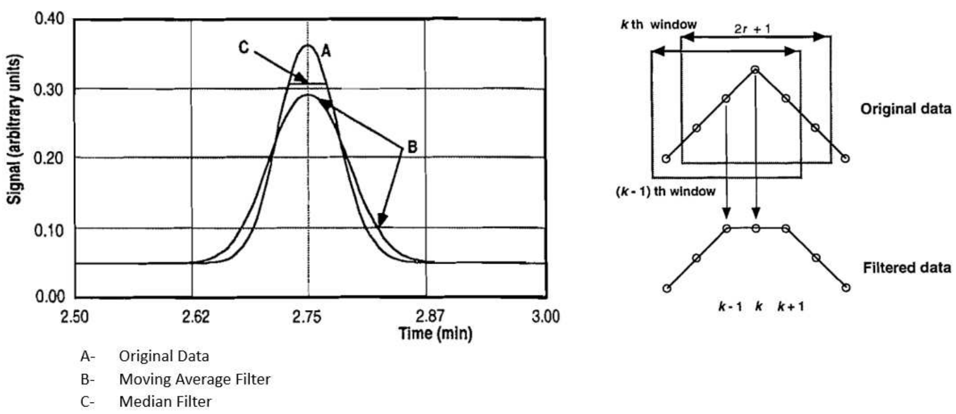


Figure 12: The Median Filter

This filter ends up gaining some advantage when compared to other filters as it manages to remove extreme fluctuations while guaranteeing the details and structure of the data. The fact is that statistically, the median can estimate the center of the data better than the average. This is why the MED filter is so robust. Nevertheless, it produces a very noisy trend that might fail the test of either smoothness or persistence.

4.3 AR(1) Model

The autoregressive model depends only on past data. Mathematically, it can be described by the following equation:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t \tag{7}$$

where the ϵ_t represents a white noise.

Often, we can forecast a series based only on the past values in the series - called lags. A time series described as a linear function of p past values with some error is called an autoregressive process of order p, and it is noted as AR(p). In a simpler scenario, the AR(1) model is an autoregressive model of order 1; each means that it only depends on one lag in the past.

$$y_t = c + \phi y_{t-1} + \epsilon_t \tag{8}$$

These time series models are long memory models, meaning they constantly factor in the previous values in the series. It can be translated mathematically in the following way: in equation 8, we have that y_t depends on y_{t-1}, so following the same logic, y_{t-1} depends on y_{t-2},

$$y_{t-1} = c + \phi y_{t-2} + \epsilon_{t-1} \tag{9}$$

$$y_t = c + \phi(c + \phi y_{t-2} + \epsilon_{t-1}) + \epsilon_t = c^* + \phi^2 y_{t-2} + \phi\epsilon_{t-1} + \epsilon_t, \tag{10}$$

where the c^* represents, a constant value. Therefore, if we continue, y_{t-2} also depends on y_{t-3}. This recursion in time goes back until the beginning of the series, until we get

$$y_t = \frac{c}{1-\phi} + \phi^t y_1 + \phi^{t-1}\epsilon_2 + \phi^{t-2}\epsilon_3 + \ldots + \epsilon_t \tag{11}$$

By understanding the equation (11), we can conclude that if the $|\phi| < 1$, the effect of shocks that happened long ago still has a little impact on the present. By implementing this condition to the parameter, we are restricting the AR(1) to stationary data, where the dependence of previous observations declines over time.

4.4 The Theta Model

The Theta Model, introduced by Assimakopoulos and Nikolopoulos in 2000, is a very accurate and robust forecasting tool for univariate time series data. Its performance in the M3-competition, the biggest time series forecasting competition, attracted the attention of researchers. Despite this simplicity, the Theta Model demonstrates to produce exceptionally accurate forecasts. (Nikolopoulos et al., 2011).

Traditionally, many forecasting methods are based on the decomposition time series data into different components, such as the seasonality, trend cycle and the irregular component. However, the Theta model is an exception, offering a unique perspective on decomposition. Assimakopoulos and Nikolopoulos defend that it is not efficiently extracting all the information from a time series by using one single method. To avoid this limitation, the Theta Model combine different lines with different values of theta, named Theta-lines, where each of these lines represents a distinct temporal perspective (short and long term components) with distinct local curvatures. (Assimakopoulos and Nikolopoulos, 2000)

According to Hyndman and Billah (2003), **Theta-lines** can be expressed as in the following way

$$Y_{t,\theta} = \hat{a}_\theta + \hat{b}_\theta(t-1) + \theta Y_t, \tag{12}$$

where $\hat{a}_\theta$ and $\hat{b}_\theta$ are the parameters of the fitted linear time trend. In a general way, we can observe the following:

- When $\theta = 0$, the decomposition reduces to a linear regression line.
- When $\theta = 1$, the series remains unchanged, representing the original series.
- When $\theta > 1$, there are increases in the local curvatures of the curves.

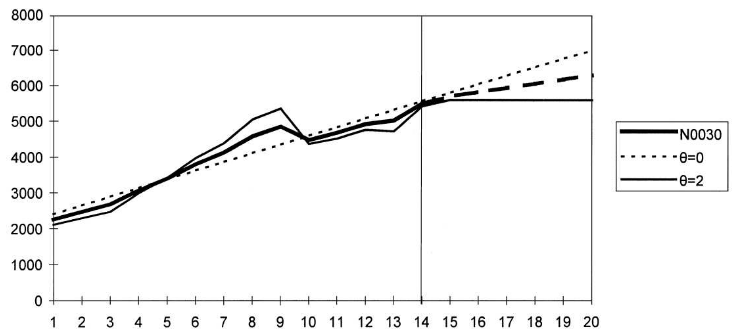


Figure 13: Theta line slope analysis

The standard Theta model starts by estimating the regression for $\theta = 0$ and $\theta = 2$:

$$(1-\theta)Y_t = \hat{a}_\theta + \hat{b}_\theta(t-1) \tag{13}$$

By extrapolating the linear time trend, we obtain:

$$\widehat{Y}_{t,0}(h) = \hat{a}_0 + \hat{b}_0(t+h-1) \tag{14}$$

In addition, by doing the SES, simple exponential smoothing, we obtain the following:

$$\widehat{Y}_{t,2}(h) = \gamma Y_{t,2} + (1-\gamma)\widehat{Y}_{t-1,2} \tag{15}$$

Note that in this specific case, $\widehat{y}_{1,2} = y_{1,2}$ and the γ is chosen to minimize the OOS error.

Then, we can already make the average between the 14 and 15 equation, we end up with the following equation:

$$\widehat{Y}_t(h) = \frac{\widehat{Y}_{t,0}(h) + \widehat{Y}_{t,2}(h)}{2} \tag{16}$$

5 Methodology

In the previous chapters, I discussed the impact of product returns on company losses and gave some insights into the data, which helped me to decide which forecast model I would study.

It is important to note that there are many different forecasting methods and models available, as I presented before, and it is crucial to carefully evaluate which is best suited for the specific context and data being analyzed. Therefore, in this chapter, I will explain the methodology that leaves me to choose one particular model by introducing the measures I used to determine which model performs best and other techniques.

5.1 Forecasting workflow

The process of applying a forecast model to our data can be characterized into some steps, as we can see in Figure 14. (Hyndman and Athanasopoulos, 2021)

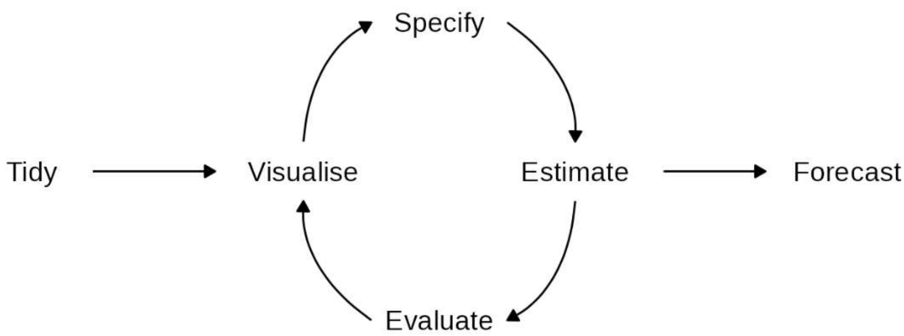


Figure 14: The Forecast Workflow

In the dataset chapter, we already tidied and visualized our data. In the models chapter, I specify what models I will use. To do the estimation, I will use cross-validation for time series, and finally, I will evaluate the performance of each model by using some metrics. This process will give me one specific model I will implement and use to forecast and drive some findings and draw conclusions.

5.2 Cross-Validation for time series

Cross-validation is a technique that helps us determine which model performs best. It is a widespread practice in machine learning. However, when dealing with time series, it can be more complex. In ML, we apply the train-test split in which we divide the data into train and test randomly. In this case, we cannot choose random samples; the way we do this division is very similar to a sliding time window that goes through the time series until reaching the end of the data, as shown in Figure 15.

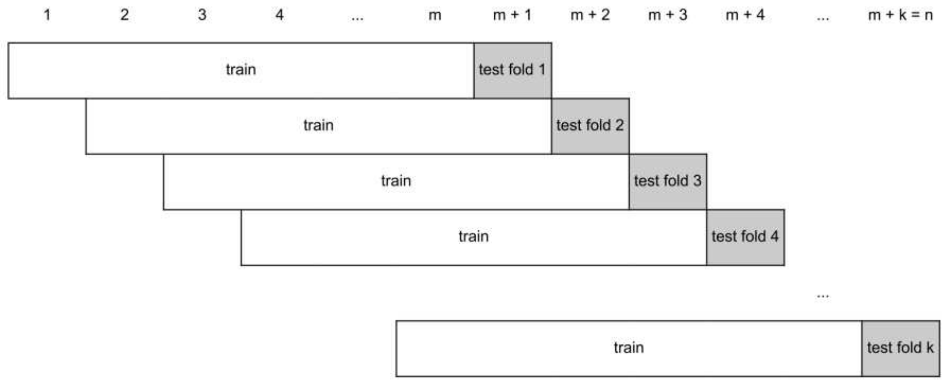


Figure 15: Cross-Validation for time series

Figure 15 illustrates how the data is divided into folds for the prediction model. The first fold comprises the initial m data points from the time series, which are utilized to train the initial prediction model. This model is then used to predict the next point $(m + 1)$. This process is repeated for each one of the K folds. By the end of the process, each fold is used as a test set once. This ensures that the algorithm learns from different periods of the series in each fold. Additionally, in each subsequent iteration, the predicted month from the previous iteration is included but with the actual value, not the predicted value. This methodology helps ensure the prediction model is accurate and reliable for future data.

5.3 Forecast accuracy

To evaluate the forecast accuracy of our models, we have several metrics where we compare predictions to the testing set. Mathematically, we can represent as the following equation,

$$e_{T+h|T} = Y_{T+h|T} - \hat{Y}_{T+h|T} \tag{17}$$

Equation (17) can be named the forecast errors or the predictions. By interpreting these measures, we can choose the best model to implement the forecast.

5.3.1 MAE - Mean Absolute Error

The mean absolute error is a prevalent metric; the following equation characterizes it:

$$MAE = \frac{1}{H} \sum_{h} |e_{T+h|T}| \tag{18}$$

What MAE does is simply the mean of absolute differences between the actual and the predictors' values.

5.3.2 MSE - Mean squared error

The mean squared error is also a widespread measure. As we can see in the following equation:

$$MSE = \frac{1}{H} \sum_{h} e^2_{T+h|T} \tag{19}$$

The MSE is the average of the squares of the errors.

5.3.3 RMSE - Root mean squared error

Root mean squared error measures how spread out the residuals are. Mathematically, it is simply the square of the mean squared error, as we can see in the following equation:

$$RMSE = \sqrt{\frac{1}{H} \sum_{h} e^2_{T+h|T}} \tag{20}$$

5.3.4 AIC - Akaike information criterion

The AIC is one measure very interesting in time series. It analyzes the model's fit to the training data and incorporates a penalization term for the complexity of the model. Mathematically, we have the following equation.

$$AIC = ln\left(\frac{e'e}{K}\right) + \frac{2q}{K}, \tag{21}$$

where K is the number of folds, q is the number of parameters of each forecasting model, and e is the column of the OOS errors.

Since all these equations measure the errors, the lower the error, the better the model. By comparing all the errors of the different models, I will conclude the best model to implement the forecast.

6 Findings

In this chapter, the main objective is to explore the results and findings I obtained by applying all the methodology explained in the previous chapter to my dataset.

The performance of these models can vary depending on the specific data. It's important to evaluate the reliability of each model before making any conclusion based on their predictions. Therefore, I decided to do the forecast for my three adjusted variables independently.

6.1 Product return rate - return_r_sa

Firstly, I tested which number of folds K provides the lowest forecast errors. Thus, the best value of K is 12. In Table 3, we can see the various results of the measures for this K.

Forecast accuracy	HP filter	MED filter	Ar(1) model	Theta model
MAE	0.3151	0.3143	0.3165	**0.2917**
MSE	0.2377	0.2351	0.2335	**0.2136**
RMSE	0.4876	0.4849	0.4832	**0.4622**
AIC	-1.27	-1.2809	-1.2879	**-1.377**

Table 3: Performance of various models for the adjusted product return rate (%)

As I mentioned, the lower the error value, the better the model performs. As a result, the best model to forecast the product return rate is the Theta model.

By implementing the Theta model, I obtained the predicted values, as well as the forecast plot for the next 12 months.

The grey area in Figure 16 represents the forecast confidence interval and is obtained using the exponential smoothing method underlying the Theta model. On the other hand, the width of the confidence interval refers to the desired level of confidence; in this case, the level of confidence is 80% and 95%.

By interpreting both Figure 16 and Table 4, we can see that the tendency of the product return in the next year will increase in the next year.

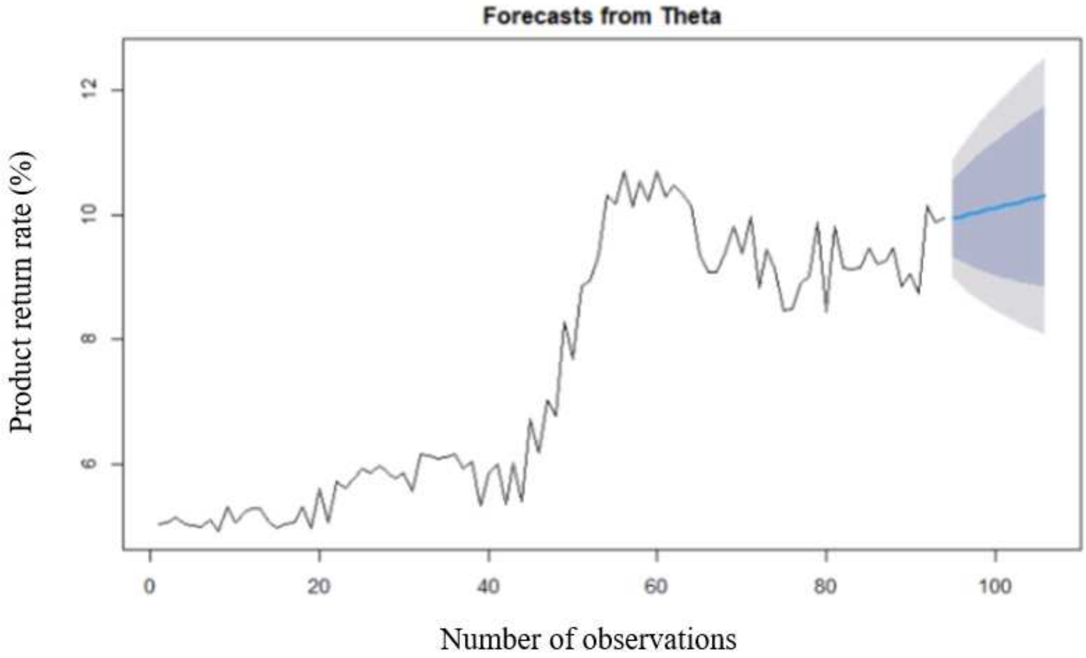


Figure 16: Forecast from Theta model for adjusted product return rate (%)

Date	**Forecast values (%)**
Nov 2023	9.952632
Dec 2023	9.984981
Jan 2024	10.017329
Feb 2024	10.049677
Mar 2024	10.082026
Apr 2024	10.114374
May 2024	10.146723
Jun 2024	10.179071
Jul 2024	10.211419
Aug 2024	10.243768
Sep 2024	10.276116
Out 2024	10.308465

Table 4: Forecasting values for the adjusted product return rate (%) next year

6.2 Cost return - cost_ret_sa

Similarly to the time series before, I tested the validation set and decided to attribute K equal to 24 since it is the value that gives me lower error values. In the Table, we can see which model to choose to make the predictions.

Forecast accuracy	HP filter	MED filter	Ar(1) model	Theta model
MAE	0.3641	0.3603	**0.356**	0.3568
MSE	0.2604	0.2543	0.2442	**0.2392**
RMSE	0.5103	0.5043	0.4942	**0.4891**
AIC	-1.2622	-1.2857	-1.3263	**-1.3471**

Table 5: Performance of various models for the adjusted company's cost (K euros)

By interpreting Table 5, although the MAE value is lower for the AR(1) model, the model with the lowest errors is the Theta model for all the other measures. Additionally, the difference between the AR(1) model's MAE value and the theta model's value is very small. Therefore, the Theta Model is the model I will apply to the cost_ret_sa variable for forecasting.

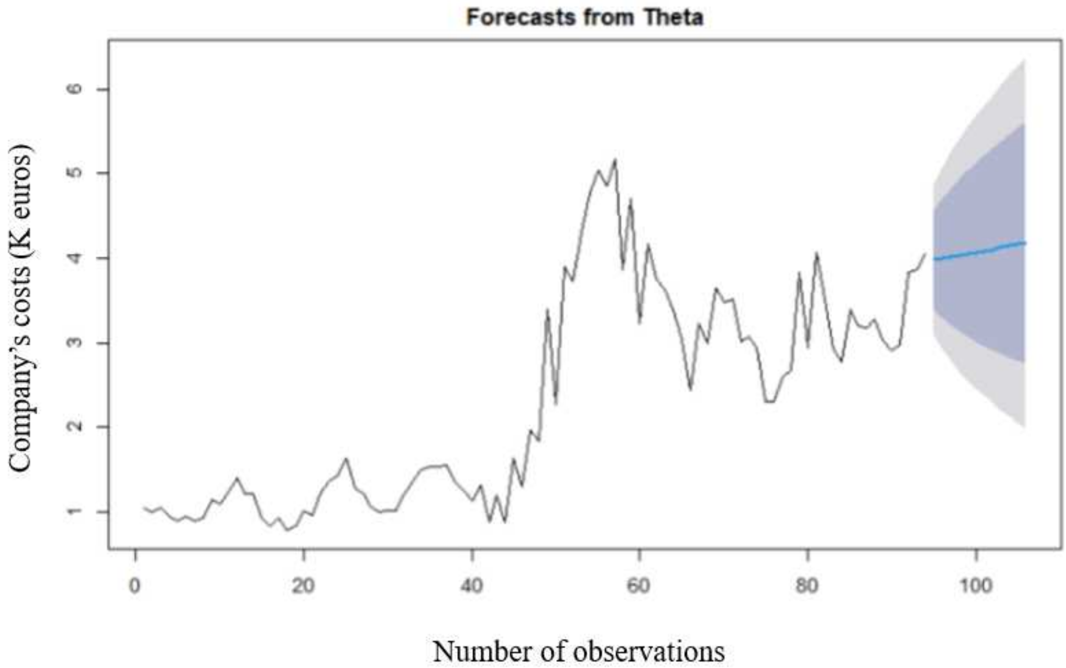


Figure 17: Forecast from Theta model for adjusted company's cost (K euros)

We can see that the company's costs will significantly increase in the next year by interpreting Figure 17 and Table 6.

Date	Forecast values (K euros)
Nov 2023	3.982960
Dec 2023	4.000653
Jan 2024	4.018346
Feb 2024	4.036039
Mar 2024	4.053731
Apr 2024	4.071424
May 2024	4.089117
Jun 2024	4.106810
Jul 2024	4.124503
Aug 2024	4.142196
Sep 2024	4.159889
Out 2024	4.177582

Table 6: Forecasting values for the adjusted company's costs (K euros) next year

6.3 Profit - profit_sa

This time series presented lower error values when K is equal to 12. In Table 7, we can understand the best model to perform better predictions in the next year.

Forecast accuracy	HP filter	MED filter	Ar(1) model	Theta model
MAE	3.2736	3.2083	3.1973	**3.077**
MSE	21.6055	20.6059	**19.8775**	20.6129
RMSE	4.6482	4.5394	**4.4581**	4.5401
AIC	3.2396	3.1922	**3.1563**	3.1926

Table 7: Performance of various models for the adjusted company's profit (K euros)

Since the MAE is the less sensitive measure to outliers, I will apply the Theta model in this case since this value is much lower. Figure 18 shows the adjusted company's profit forecast in K euros.

By interpreting Figure 18, we can see a slight decrease in the company's profit next year. Table 8 shows us the predictions for each month from November 2023 to October 2024.

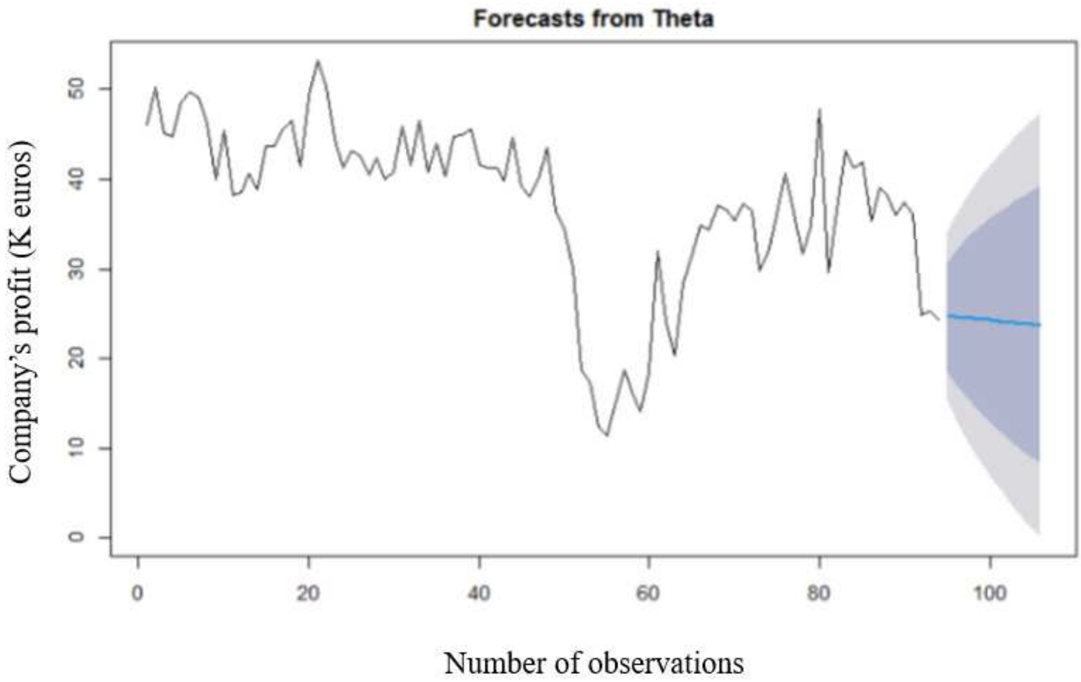


Figure 18: Forecast from Theta model for adjusted company's profit (K euros)

Date	Forecast values (K euros)
Nov 2023	24.70351
Dec 2023	24.61448
Jan 2024	24.52545
Feb 2024	24.43642
Mar 2024	24.34739
Apr 2024	24.25835
May 2024	24.16932
Jun 2024	24.08029
Jul 2024	23.99126
Aug 2024	23.90223
Sep 2024	23.81320
Out 2024	23.72417

Table 8: Forecasting values for the adjusted company's profit (K euros) next year

7 Discussion

In this chapter, the main objective is to discuss the findings we concluded in the previous chapter and understand the limitations and constraints.

As mentioned in the literature and review, consumer satisfaction and behavior significantly impact product returns. Therefore, although we can make certain conclusions and gain some insights about what will happen next year by applying the Theta Model, the results we obtain approximate what will happen if the company doesn't use any strategy to avoid the impact of product return in their business.

In addition, by only having three variables in the study, I cannot numerically quantify the number or a percentage that tells us that applying the forecast as a mitigation strategy will have an X increase in the company's profit. To achieve that goal, I would need much more information.

Other limitation it that I only tested four different models, which cannot be sufficient, and it might occur that other models have better predictions and lower error values.

Finally, these results are obtained for analyzing one specific fashion e-commerce company, so we can not conclude these results in a general way, for other types of companies and sectors.

8 Conclusions

In this chapter, I will resume the main findings and present future research. One of the primary purposes of my thesis is to understand if the amount of product returns has a direct impact on a company's losses, and this is verified, i.e., the higher the product return rate, the higher the company's losses, as well as the higher both variables, the lower the companies profit. In addition, when comparing the forecast models, we concluded that the Theta Model has better performance, which will result in more realistic predictions.

Since we already know that if the company doesn't change its strategy, its profit will decrease, the next step of this research could be understanding what if the company applies some approach. Analyzing and comparing the results I obtained with the values after applying some strategy to reduce the impact of product returns will be a better way to understand the importance of using forecasts to manage such implications. On the other hand, continuous study will provide more information about the consumer's behavior, making it easier to find the best strategy to overcome such problems.

According to Petersen and Kumar (2015), companies can gain a real competitive advantage if they consider both strategies of minimizing product returns and understanding consumer behavior. As a result, if the company redefines processes and uses forecast models to manage and to know precisely when to apply each strategy, it will be more able to reduce the likelihood of product returns and improve its overall profitability. By leveraging this knowledge, businesses can stay ahead of the competition and maintain a strong market position.

Another type of research that can be made is to study the variable product complaints and understand the relationship between this variable and the product return rate, company's profits and losses. This way, we will better understand how consumer satisfaction influences the amount of product returns. As we saw by interpreting the predictions, the tendency is for an increase in the number of returns from consumers, which generates an increase in costs and, consequently, a decrease in the company's profits. This might be because consumers are unhappy with the online service or the quality of the products, so studying variable product complaints is essential.

Another way to approach my research question is using machine learning: if I had more variables, for example, the sales of each product, we could accurately predict consumer demand and preferences; companies can optimize their inventory and ensure that they only produce and sell products in high order. This approach reduces the number of returns and enhances the overall customer experience and brand loyalty.

Another area for improvement is that I only studied four different models - HP filter, MED filter, AR(1) model, and Theta model, as there are many other models to forecast. As a result, although between these four models, the one that performs the best is the Theta model, it might happen that other models that I didn't study perform better forecasts. Therefore, this can be a good idea to apply to future research.

To conclude, it is crucial to control the number of product returns periodically and devise plans and strategies to minimize their impact since it is proved in this specific case that it might lead to a massive increase in company losses. Continuously controlling will help companies make informed decisions, and they will be able to understand their consumers better and apply the best strategies by using forecast models and machine learning techniques.

References

J. Ader, P. Adhi, J. Chai, M. Singer, S. Touse, and H. Yankelevich. Returning to order: Improving returns management for apparel companies. McKinsey's Consumer and Retail Practice, 5 2021. URL https://www.mckinsey.com/industries/retail/our-insights/returning-to-order-improving-returns-management-for-apparel-companies#/.

S. Agrawal and R. K. Singh. Forecasting product returns and reverse logistics performance: structural equation modelling. Management of Environmental Quality: An International Journal, 31:1223–1237, 10 2019. ISSN 1477-7835. doi: 10.1108/MEQ-05-2019-0109.

V. Assimakopoulos and K. Nikolopoulos. The theta model: a decomposition approach to forecasting. International Journal of Forecasting, 16:521–530, 10 2000. ISSN 01692070. doi: 10.1016/S0169-2070(00)00066-2.

N. N. Bechwati and W. S. Siegal. The impact of the prechoice process on product returns. Journal of Marketing Research, 42:358–367, 8 2005. ISSN 0022-2437. doi: 10.1509/jmkr.2005.42.3.358.

J. Calma. Free returns come with an environmental cost. 2019. URL https://www.theverge.com/2019/12/26/21031855/free-returns-environmental-cost-holiday-online-shopping-amazon.

S. Chevalier. E-commerce returns in europe - statistics facts. Statista, 2023. URL https://www.statista.com/topics/11361/e-commerce-returns-in-europe/#topicOverview.

T. Doering and N. C. Suresh. Forecasting and performance: Conceptualizing forecasting management competence as a higher-order construct. Journal of Supply Chain Management, 52:77–91, 10 2016. ISSN 15232409. doi: 10.1111/jscm.12111. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jscm.12111.

C. DUHS. The challenge of e-commerce returns post covid-19 pandemic. 2020. URL https://www.360idtag.com/blogs/news/challenge-of-ecommerce-returns-post-covid-19-pandemic.

J. Eklof, O. Podkorytova, and A. Malova. Linking customer satisfaction with financial performance: an empirical study of scandinavian banks. Total Quality Management Business Excellence, 31:1684–1702, 11 2020. ISSN 1478-3363. doi: 10.1080/14783363.2018.1504621.

W. Enders, J. Wiley, and Sons;. Applied econometric time series. Applied Econometric Time Series, 1995.

C. Fornell, M. D. Johnson, E. W. Anderson, J. Cha, and B. E. Bryant. The american customer satisfaction index: Nature, purpose, and findings. Journal of Marketing, 60: 7–18, 10 1996. ISSN 0022-2429. doi: 10.1177/002224299606000403.

R. Hadiantini, Silalahi, and H. Hendrayati. The importance of consumer satisfaction for e-commerce users: How that affects consumer behavior on consumer satisfaction? 2021. doi: 10.2991/assehr.k.210312.016.

J. D. Hamilton. Why you should never use the hodrick-prescott filter. The Review of Economics and Statistics, 100:831–843, 12 2018. ISSN 0034-6535. doi: 10.1162/rest_a_00706.

E. Higueras-Castillo, F. J. Liébana-Cabanillas, and Ángel F. Villarejo-Ramos. Intention to use e-commerce vs physical shopping. difference between consumers in the post-covid era. Journal of Business Research, 157:113622, 3 2023. ISSN 01482963. doi: 10.1016/j.jbusres.2022.113622. URL https://www.sciencedirect.com/science/article/pii/S0148296322010876.

R. J. Hodrick and E. C. Prescott. Postwar u.s. business cycles: An empirical investigation. Journal of Money, Credit and Banking, 29:1, 2 1997. ISSN 00222879. doi: 10.2307/2953682.

R. J. Hyndman and G. Athanasopoulos. Forecasting: Principles and Practice. Monash University, Australia, 3rd edition, 2021. URL https://otexts.com/fpp3/.

R. J. Hyndman and B. Billah. Unmasking the theta method. International Journal of Forecasting, 19:287–290, 4 2003. ISSN 01692070. doi: 10.1016/S0169-2070(01)00143-1. URL https://www.sciencedirect.com/science/article/abs/pii/S0169207001001431.

W. Kammerer, C. Bimschleger, and M. Leddy. Deloitte study. 2020.

C. Lee and C. K. Morewedge. Mental accounting of product returns. Journal of Consumer Psychology, 33:583–590, 7 2023. ISSN 1057-7408. doi: 10.1002/jcpy.1354. URL https://myscp.onlinelibrary.wiley.com/doi/full/10.1002/jcpy.1354.

M. McFall-Johnsen. These facts show how unsustainable the fashion industry is. World Economic Forum, 2020. URL https://www.weforum.org/agenda/2020/01/fashion-industry-carbon-unsustainable-environment-pollution/.

A. C. Neiva. Marketing strategies shein. 2023. URL https://edrone.me/pt/blog/shein-crescimento-a-qualquer-custo.

K. Nikolopoulos, V. Assimakopoulos, N. Bougioukos, A. Litsa, and F. Petropoulos. The Theta Model: An Essential Forecasting Tool for Supply Chain Planning, pages 431–437. 2011. doi: 10.1007/978-3-642-25646-2_56.

NRF. Customer returns in the retail industry. National Retail Federation, 2021. URL https://nrf.com/research/customer-returns-retail-industry.

A. Paswan. Consumers' legitimate and opportunistic product return behaviors in online shopping. Journal of Electronic Commerce Research, 2018.

J. A. Petersen and V. Kumar. Are product returns a necessary evil? antecedents and consequences. Journal of Marketing, 73:35–51, 5 2009. ISSN 0022-2429. doi: 10.1509/jmkg.73.3.035.

J. A. Petersen and V. Kumar. Can product returns make you money? 7, 2010.

J. A. Petersen and V. Kumar. Perceived risk, product returns, and optimal resource allocation: Evidence from a field experiment. Journal of Marketing Research, 52:268–285, 4 2015. ISSN 0022-2437. doi: 10.1509/jmr.14.0174. URL https://journals.sagepub.com/doi/10.1509/jmr.14.0174.

F. Petropoulos, D. Apiletti, V. Assimakopoulos, M. Z. Babai, D. K. Barrow, S. B. Taieb, C. Bergmeir, R. J. Bessa, J. Bijak, J. E. Boylan, J. Browell, C. Carnevale, J. L. Castle, P. Cirillo, M. P. Clements, C. Cordeiro, F. L. C. Oliveira, S. D. Baets, A. Dokumentov, J. Ellison, P. Fiszeder, P. H. Franses, D. T. Frazier, M. Gilliland, M. S. Gönül, P. Goodwin, L. Grossi, Y. Grushka-Cockayne, M. Guidolin, M. Guidolin, U. Gunter, X. Guo, R. Guseo, N. Harvey, D. F. Hendry, R. Hollyman, T. Januschowski, J. Jeon, V. R. R. Jose, Y. Kang, A. B. Koehler, S. Kolassa, N. Kourentzes, S. Leva, F. Li, K. Litsiou, S. Makridakis, G. M. Martin, A. B. Martinez, S. Meeran, T. Modis, K. Nikolopoulos, D. Önkal, A. Paccagnini, A. Panagiotelis, I. Panapakidis, J. M. Pavía, M. Pedio, D. J. Pedregal, P. Pinson, P. Ramos, D. E. Rapach, J. J. Reade, B. Rostami-Tabar, M. Rubaszek, G. Sermpinis, H. L. Shang, E. Spiliotis, A. A. Syntetos, P. D. Talagala, T. S. Talagala, L. Tashman, D. Thomakos, T. Thorarinsdottir, E. Todini, J. R. T. Arenas, X. Wang, R. L. Winkler, A. Yusupova, and F. Ziel. Forecasting: theory and practice. International Journal of Forecasting, 38:705–871, 7 2022. ISSN 01692070. doi: 10.1016/j.ijforecast.2021.11.001. URL https://www.sciencedirect.com/science/article/pii/S0169207021001758.

A. M. Ramírez. Product return and logistics knowledge: Influence on performance of the firm. Transportation Research Part E: Logistics and Transportation Review, 48: 1137–1151, 11 2012. ISSN 13665545. doi: 10.1016/j.tre.2012.06.001.

W. J. Reinartz and V. Kumar. The impact of customer relationship characteristics on profitable lifetime duration. Journal of Marketing, 67:77–99, 1 2003. ISSN 0022-2429. doi: 10.1509/jmkg.67.1.77.18589.

M. Rossi. What is wardrobing and how do you prevent it from happening? narvar, 2022. URL https://corp.narvar.com/blog/how-to-prevent-wardrobing.

J. SCHIFFER. The unsustainable cost of free returns. VOGUE, 2019. URL https://www.voguebusiness.com/consumers/returns-rising-costs-retail-environmental.

Schiffman, L. G, Wisenbilt, Joseph, Kumar, and S. Ramesh. Consumer behaviour book. Pearson India Education Services Pvt. Ltd., 2014. URL https://books.google.pt/books?hl=pt-PT&lr=&id=VDXiBAAAQBAJ&oi=fnd&pg=PP1&ots=csgfduiP9q&sig=V5UHIWSYPZUBV0wRANMT5WHBCpY&redir_esc=y#v=onepage&q&f=false.

J. D. Shulman, A. T. Coughlan, and R. C. Savaskan. Optimal restocking fees and information provision in an integrated demand-supply model of product returns. Manufacturing Service Operations Management, 11:577–594, 10 2009. ISSN 1523-4614. doi: 10.1287/msom.1090.0256.

P. Silayoi and M. Speece. The importance of packaging attributes: a conjoint analysis approach. European Journal of Marketing, 41:1495–1517, 11 2007. ISSN 0309-0566. doi: 10.1108/03090560710821279.

P. St-Amant and S. van Norden. Measurement of the output gap: A discussion of recent research at the bank of canada. Technical Report No. 79 (English), 8 1997. URL https://www.bankofcanada.ca/1997/08/technical-report-no79/.

Statista. Reverse logistic - statista. Journal of Reverse Logistics, 33, 2021. URL https://www.statista.com/study/48655/reverse-logistics/.

D. C. Stone. Application of median filtering to noisy data. Canadian Journal of Chemistry, 73:1573–1581, 10 1995. ISSN 0008-4042. doi: 10.1139/v95-195.

A. Sundar and T. J. Noseworthy. Place the logo high or low? using conceptual metaphors of power in packaging design. Journal of Marketing, 78:138–151, 9 2014. ISSN 0022-2429. doi: 10.1509/jm.13.0253.

G. Taher. E-commerce: Advantages and limitations. International Journal of Academic Research in Accounting, Finance and Management Sciences, 11, 2 2021. ISSN 2225-8329. doi: 10.6007/IJARAFMS/v11-i1/8987.

L. Thomas. More online sales mean retailers need to solve a 50 billion dollars returns problem this holiday season. CNBC, 2019. URL https://www.cnbc.com/2019/12/19/online-returns-this-holiday-season-to-hit-record-41point6-billion.html.

S. Thomas. 8 common reasons for product returns and ways to minimize them. avada, 2023. URL https://avada.io/resources/product-return.html#:~:text=8%20most%20common%20reasons%20for%20products%20return%20and,...%208%208.%20Changes%20in%20customers%E2%80%99%20opinion%20.

Y. Wen and B. Zeng. A simple nonlinear filter for economic time series analysis. Economics Letters, 64:151–160, 8 1999. ISSN 01651765. doi: 10.1016/S0165-1765(99)00089-0.

T. Williams. Using amazon return data to your advantage and reducing return rates. Envision Horizons, 2021. URL https://www.envisionhorizons.com/blog/using-amazon-return-data-to-your-advantage-and-reducing-return-rates.

X. Xiaofeng and F. Tijun. Forecast for the amount of returned products based on wave function. pages 324–327. IEEE, 2009. ISBN 978-0-7695-3876-1. doi: 10.1109/ICIII.2009.235. URL https://ieeexplore.ieee.org/abstract/document/5369856.

S. Yan, T. W. Archibald, X. Han, and Y. Bian. Whether to adopt "buy online and return to store" strategy in a competitive market? European Journal of Operational Research, 301:974–986, 9 2022. ISSN 03772217. doi: 10.1016/j.ejor.2021.11.040. URL https://www.sciencedirect.com/science/article/abs/pii/S0377221721009917.

D. Zhang, R. Frei, G. Wills, E. Gerding, S. Bayer, and P. K. Senyo. Strategies and practices to reduce the ecological impact of product returns: An environmental sustainability framework for multichannel retail. Business Strategy and the Environment, 2 2023. ISSN 0964-4733. doi: 10.1002/bse.3385. URL https://onlinelibrary.wiley.com/doi/full/10.1002/bse.3385.

C. Zikopoulos. Remanufacturing lotsizing with stochastic lead-time resulting from stochastic quality of returns. International Journal of Production Research, 55:1565–1587, 3 2017. ISSN 0020-7543. doi: 10.1080/00207543.2016.1150616.

CODE APPENDIX

```
library(dplyr)
library(tidyr)
library(ggplot2)
library(readr)
library(readxl)
library(tsibble)
library(fable)
library(feasts)
library(lubridate)
library(patchwork)
library(stargazer)
library(forecast)
library(skimr)
```

#Clean the environment

```
rm(list=ls())
```

#Load the small library of some methods

```
source("C:/Users/user/Documents/DISSERTATION/DADOS/librec.R")
```

#Import data

```
dt <- read.csv("C:/Users/user/Documents/DISSERTATION/DADOS/dados_final.csv", sep=";")
head(dt,4)
```

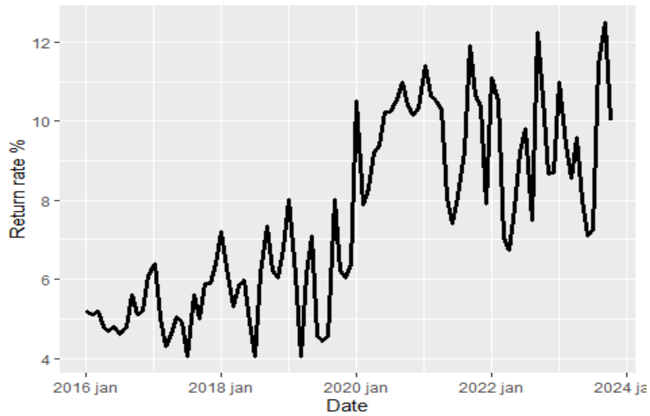
```
##      date return_rate   profit cost_return
## 1 2016-01         5.2 43.23750     1.02464
## 2 2016-02         5.1 48.71363     0.99904
## 3 2016-03         5.2 44.60363     1.06880
## 4 2016-04         4.8 43.20350     0.91200
```

#Transform in tsibble

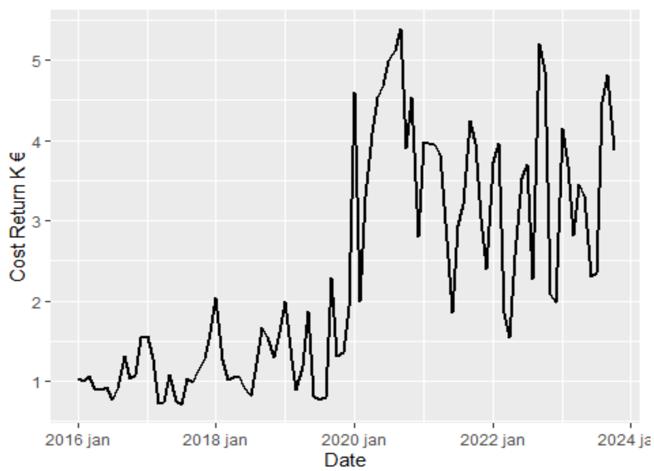
```
dts <- dt %>% mutate(date = yearmonth(date)) %>% as_tsibble(index=date)
```

#Analyze Variables

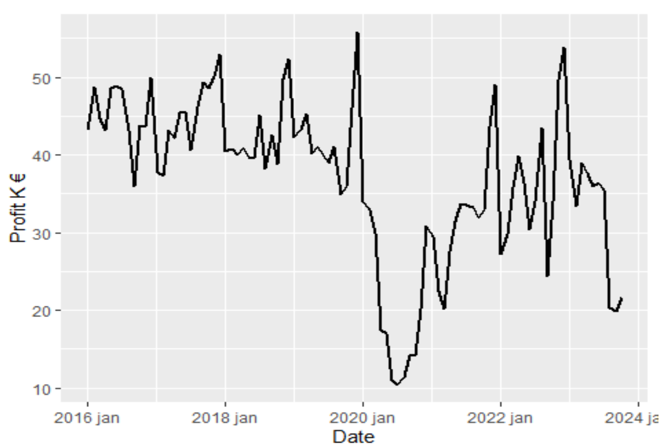
```
ggplot(dts, aes(x=date, y=return_rate)) + geom_line(size=1.2) + labs(x= "Date", y= "Return rate %")
```




```
ggplot(dts, aes(x=date, y=cost_return)) + geom_line(size=1) + labs(x= "Date", y= "Cost Return K €")
```

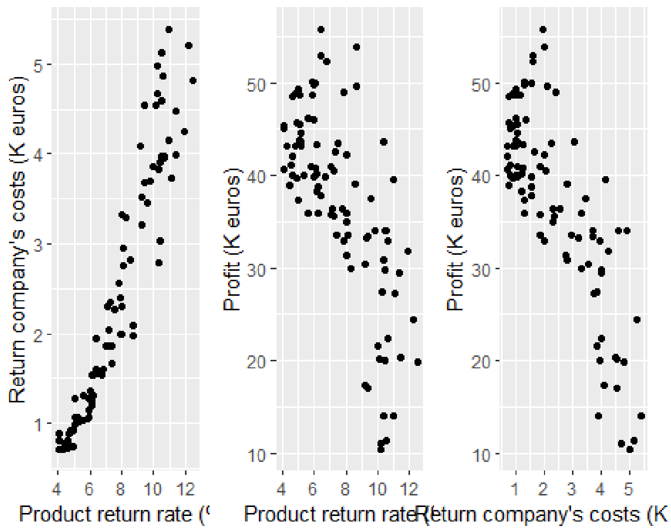



```
ggplot(dts, aes(x=date, y=profit)) + geom_line(size=1) + labs(x= "Date", y= "Profit K €")
```



#Scatter plots - relationship between variables

```
rc <- ggplot(dts, aes(x=return_rate, y=cost_return)) + geom_point() + labs(x= "Product return rate (%)", y= "Return company's costs (K euros)")
rp <- ggplot(dts, aes(x=return_rate, y=profit)) + geom_point() + labs(x= "Product return rate (%)", y= "Profit (K euros)")
cp <- ggplot(dts, aes(x=cost_return, y=profit)) + geom_point() + labs(x= "Return company's costs (K euros)", y= "Profit (K euros)")
rc + rp + cp
```



I desazonalized the time series with the Gretl application - TRAMO-SEATS.

#Import Adjusted data

```
data <- read.csv("C:/Users/user/Documents/DISSERTATION/DADOS/dadosfinal_sa.csv", sep=";")

#Transform the obs variable from 2016M01 to 2016-01
data <- data %>% mutate(obs = gsub("M", "-", obs))

skim(data)
```

Data summary

Name	data
Number of rows	94
Number of columns	4
Column type frequency:	
character	1
numeric	3

Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
obs	0	1	7	7	0	94	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
return_r_sa	0	1	7.49	2.04	4.94	5.57	6.89	9.37	10.70	▇▂▁▃▃
profit_sa	0	1	37.27	9.45	11.39	34.55	39.89	43.67	53.25	▁▁▃▇▂
cost_ret_sa	0	1	2.30	1.26	0.79	1.16	1.90	3.35	5.17	▇▁▂▁▁

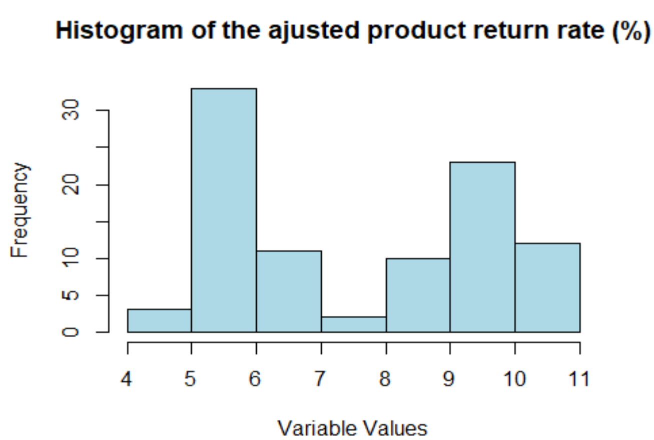
#Transform data in tsibble

```
datas <- data %>% mutate(date = yearmonth(obs)) %>% as_tsibble(index=date)
```

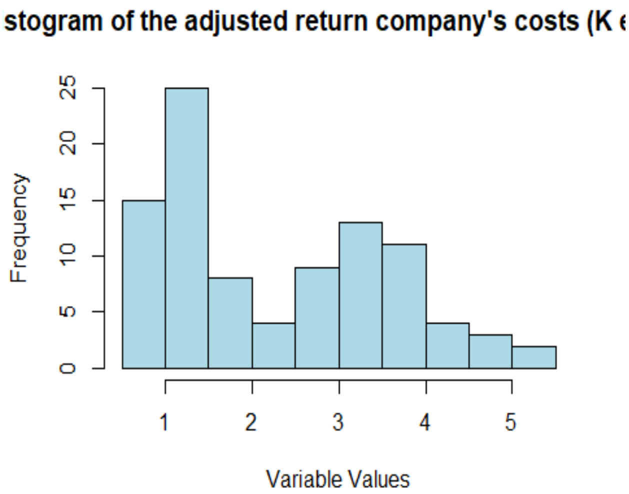
#Analysis of the adjusted time series

##Histograms

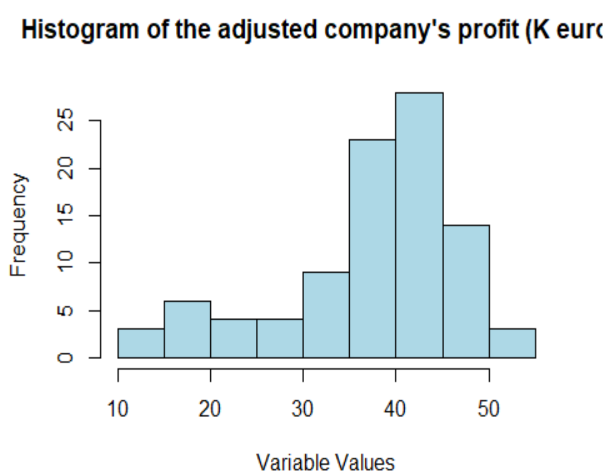
```
#Product return rate
hist(datas$return_r_sa, main = "Histogram of the ajusted product return rate (%)", xlab = "Variab
le Values", ylab = "Frequency", col = "lightblue")
```



```
#Costs
hist(datas$cost_ret_sa, main = "Histogram of the adjusted return company's costs (K euros)", xla
b = "Variable Values", ylab = "Frequency", col = "lightblue")
```

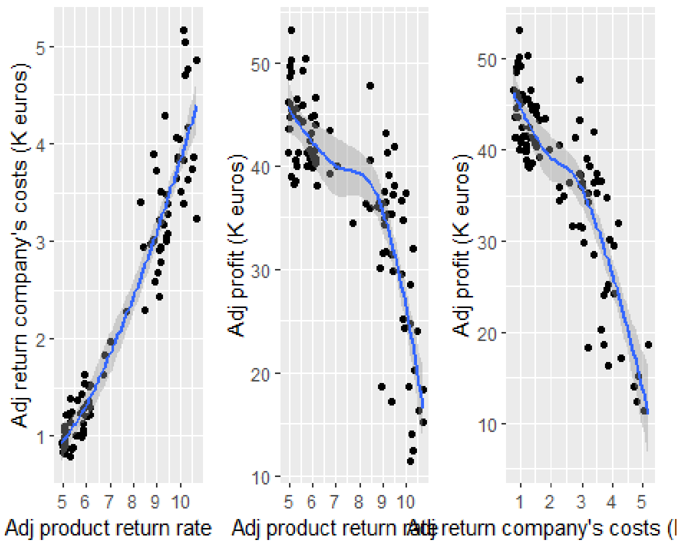



```
#Profits
hist(datas$profit_sa, main = "Histogram of the adjusted company's profit (K euros)", xlab = "Vari
able Values", ylab = "Frequency", col = "lightblue")
```



#Scatter plots - relationship between adjusted variables

```
rc2 <- ggplot(datas, aes(x=return_r_sa, y=cost_ret_sa)) + geom_point() + labs(x= "Adj product re
turn rate (%)", y= "Adj return company's costs (K euros)") + geom_smooth(method = "auto")
rp2 <- ggplot(datas, aes(x=return_r_sa, y=profit_sa)) + geom_point() + labs(x= "Adj product retu
rn rate (%)", y= "Adj profit (K euros)") + geom_smooth(method = "auto")
cp2 <- ggplot(datas, aes(x=cost_ret_sa, y=profit_sa)) + geom_point() + labs(x= "Adj return comp
any's costs (K euros)", y= "Adj profit (K euros)") + geom_smooth(method = "auto")
rc2 + rp2 + cp2
```




```
matriz <- data.frame(data$return_r_sa, data$cost_ret_sa, data$profit_sa)
matriz_correlation <- cor(matriz)
print(matriz_correlation)
```

```
##                  data.return_r_sa data.cost_ret_sa data.profit_sa
## data.return_r_sa        1.0000000        0.9533178     -0.7858996
## data.cost_ret_sa        0.9533178        1.0000000     -0.8406739
## data.profit_sa         -0.7858996       -0.8406739      1.0000000
```

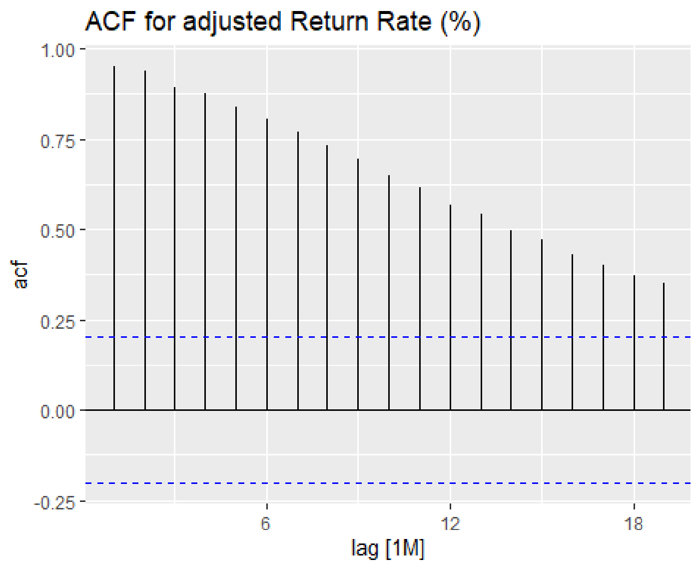
#Summary Statistics (original variables and adjusted variables)

```
stargazer(data.frame(dts, datas), type = "text", no.space = TRUE, header = FALSE , title = "Summary Statistics")
```

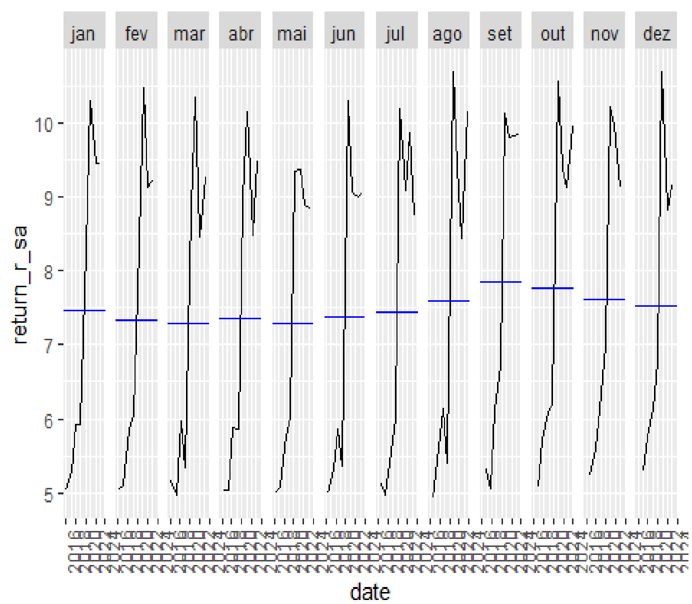
```
##
## Summary Statistics
## ===============================================
## Statistic   N   Mean  St. Dev.  Min    Max
## -----------------------------------------------
## return_rate 94 7.525   2.356   4.050  12.500
## profit      94 37.070  10.424  10.400 55.823
## cost_return 94 2.339   1.406   0.707  5.393
## return_r_sa 94 7.491   2.035   4.935  10.699
## profit_sa   94 37.273  9.453   11.389 53.254
## cost_ret_sa 94 2.298   1.262   0.786  5.173
## -----------------------------------------------
```

#Patterns analysis

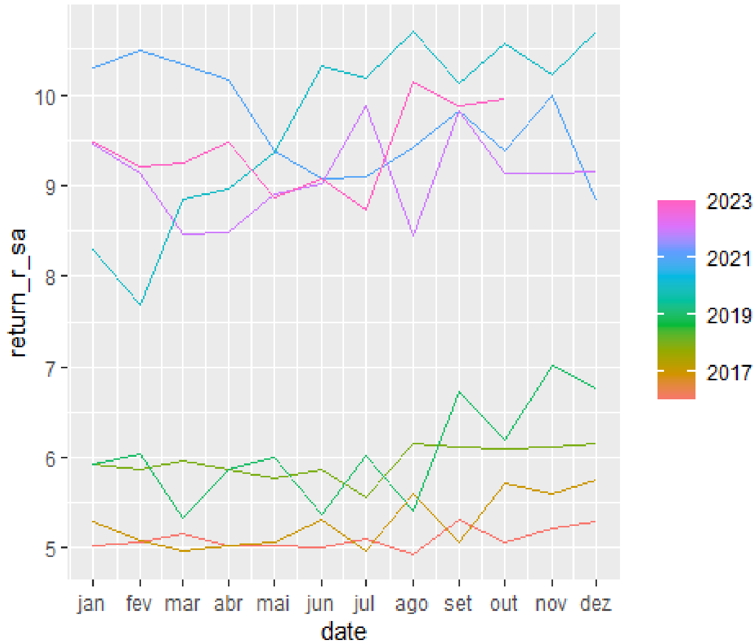
```
#Return rate (%)
datas %>% ACF(return_r_sa) %>% autoplot() + labs(title= "ACF for adjusted Return Rate (%)" )
```



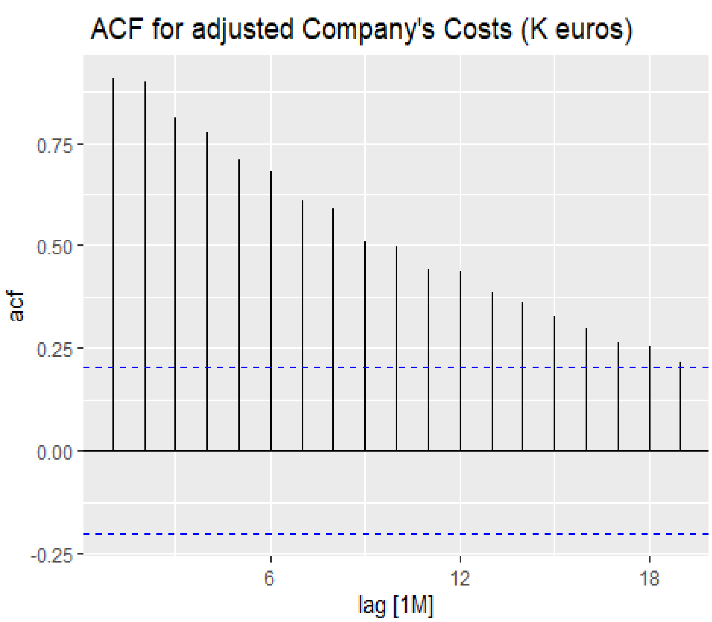

```
datas %>% gg_subseries(return_r_sa)
```



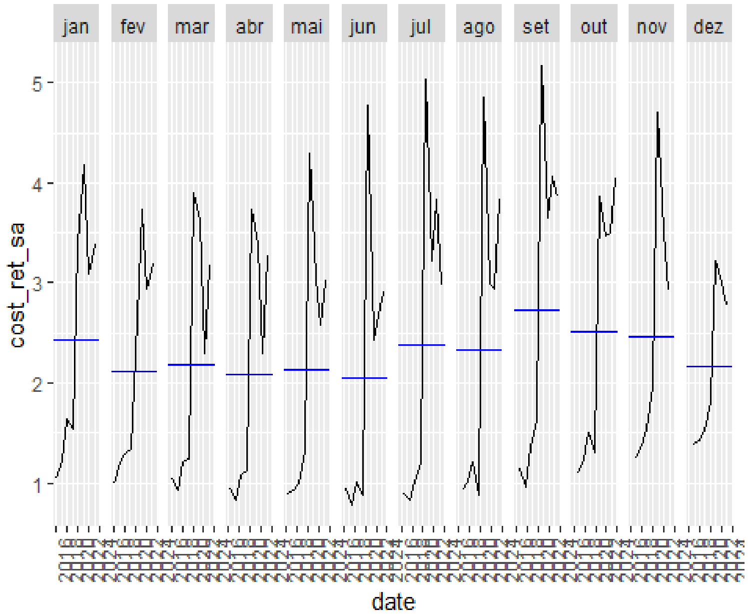

```
datas %>% gg_season(return_r_sa)
```



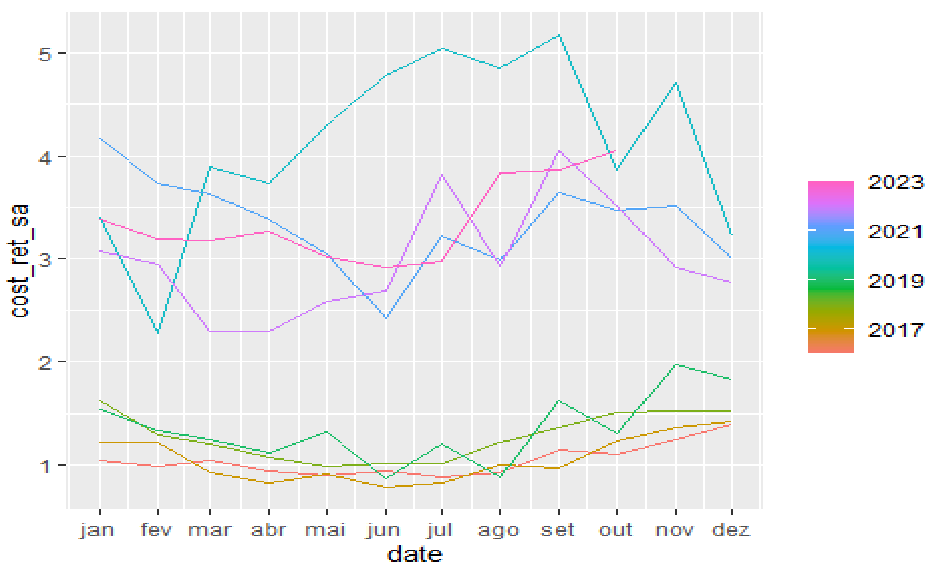

```
#Costs (K euros)
datas %>% ACF(cost_ret_sa) %>% autoplot() + labs(title=" ACF for adjusted Company's Costs (
K euros)")
```



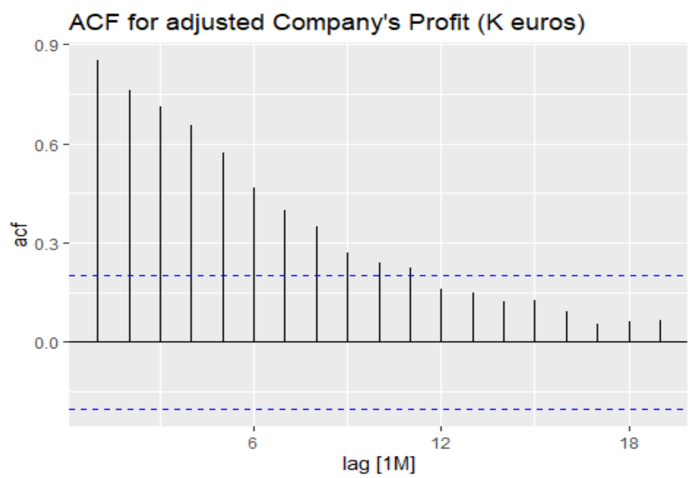

```
datas %>% gg_subseries(cost_ret_sa)
```



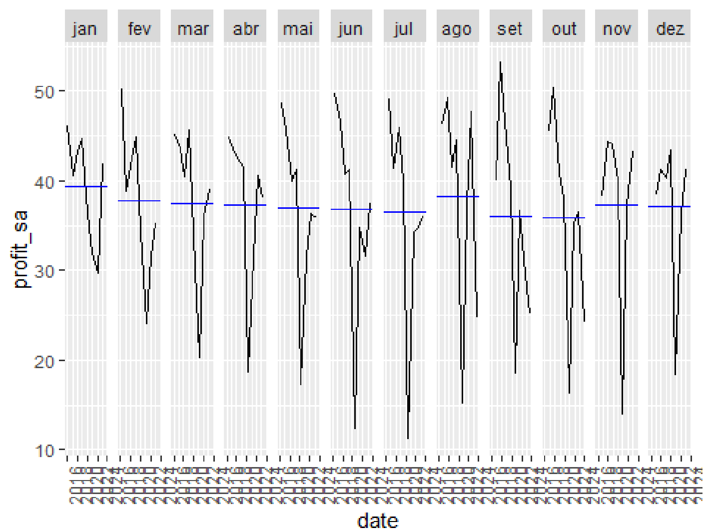

```
datas %>% gg_season(cost_ret_sa)
```



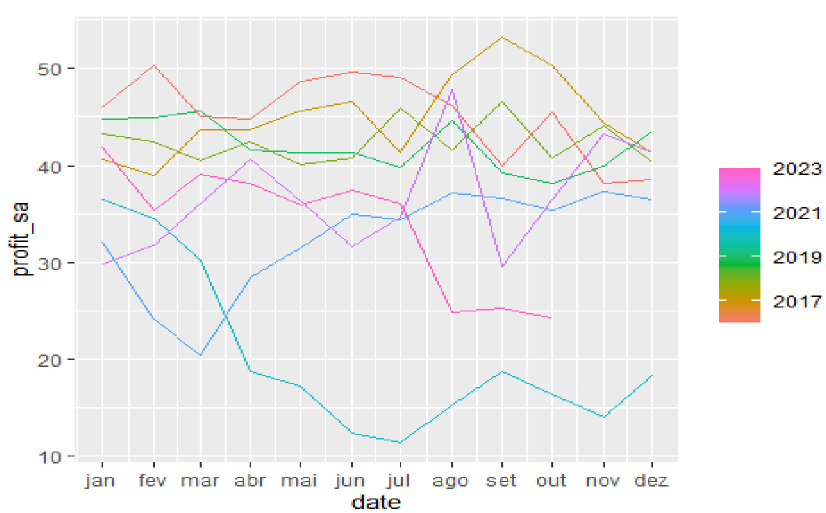

```
#Profit
datas %>% ACF(profit_sa) %>% autoplot() + labs(title="ACF for adjusted Company's Profit (K euros)")
```



datas %>% gg_subseries(profit_sa)



datas %>% gg_season(profit_sa)



##Methodology part

Here I am trying to figure our what is the best model to predict each time-series. In this case, I only illustrate one example for the product return rate. We have to repeat this process for the three different variables.

```
n <- nrow(data)
k <- 12
m <- n - k

rreturn <- data[,2]
rprofit <- data[,3]
rcost <- data[,4]
```

#Return rate sa

```
pred_AR1 <- cv.AR1(rreturn, k)   # Autoregressive benchmark model
pred_theta <- cv.theta(rreturn, k, gamma = 0.9) #Theta model
pred_hp <- cv.hp(rreturn, lambda=14400, k) #monthly data
pred_med <- cv.med(rreturn, p=35, k) #p= 36-1 - monthly data
```

Median model (simple combining procedure) - median of all models

```
P <- cbind(AR1=pred_AR1$prediction, hpfilter=pred_hp$prediction, medfilter=pred_med$predi
ction, theta=pred_theta$prediction)

pred_median <- apply(P,1,median)    # Median by row of P

error_median <- rreturn[(n-k+1):n] - pred_median

MAE_median <- sum(abs(error_median))/k
MSE_median <- sum(error_median*error_median)/k

AIC_median <- log(sum(error_median*error_median)/k) + 2/k

P <- cbind(P, median=pred_median, obs=rreturn[(n-k+1):n])
```

Race between models

```
MAE  <- cbind(hpfilter=pred_hp$MAE, medfilter=pred_med$MAE, AR1=pred_AR1$MAE, the
ta=pred_theta$MAE, median=MAE_median)
MSE  <- cbind(hpfilter=pred_hp$MSE, medfilter=pred_med$MSE, AR1=pred_AR1$MSE, theta
=pred_theta$MSE, median=MSE_median)
RMSE <- sqrt(MSE)
AIC  <- cbind(hpfilter=pred_hp$AIC, medfilter=pred_med$AIC, AR1=pred_AR1$AIC, theta=pr
ed_theta$AIC, median=AIC_median)
```

```r
#round(P,3) forecast using the methods
round(MAE,4)

##      hpfilter medfilter    AR1  theta median
## [1,]   0.3151    0.3143 0.3165 0.2917 0.3126

round(MSE,4)

##      hpfilter medfilter    AR1  theta median
## [1,]   0.2377    0.2351 0.2335 0.2136 0.2345

round(RMSE,4)

##      hpfilter medfilter    AR1  theta median
## [1,]   0.4876    0.4849 0.4832 0.4622 0.4842

round(AIC,4)

##      hpfilter medfilter     AR1  theta  median
## [1,]    -1.27   -1.2809 -1.2879 -1.377 -1.2838
```

I concluded that the Theta model is the best model to predict the time-series; the next step is implementing it.

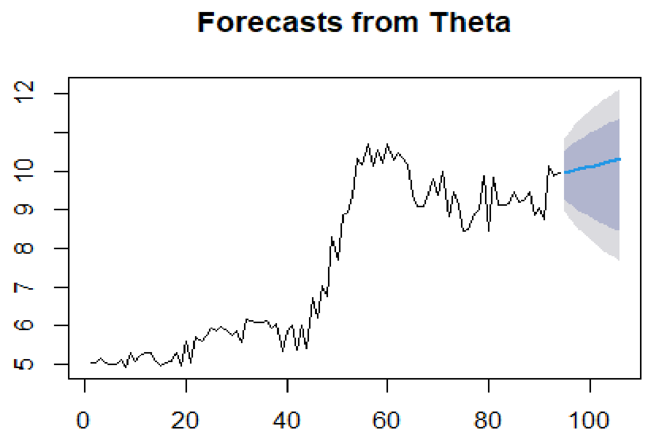
#The Theta model - Forecast

```r
thetamodel <- function(serie, h = ifelse(frequency(serie) > 1, 2 * frequency(serie), 12),
                       level = c(80, 95), x = serie) {

  # Seasonality check
  x <- as.ts(x)
  if (frequency(x) > 1 && !is.constant(x) && length(x) > 2 * frequency(x)) {
    t <- as.numeric(acf(x, lag.max = frequency(x), plot = FALSE)$acf)[-1]
    statistic <- sqrt((1 + 2 * sum(t[-frequency(x)]^2)) / length(x))
    seasonality <- (abs(t[frequency(x)]) / statistic > qnorm(0.95))
  }
  else {
    seasonality <- FALSE
  }

  # Decomposition - seasonality
  originalx <- x
  if (seasonality) {
  decomp <- decompose(x, type = "multiplicative")
    if (any(abs(seasonality(decomp)) < 1e-4)) {
      warning("Theta model (non-seasonal)")
    } else {
   x <- seasadj(decomp)
    }
  }
```

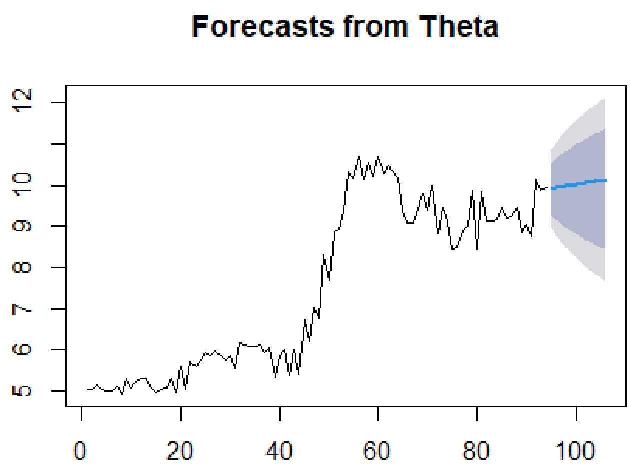
```
  # Theta lines
  forecast <- ses(datas$return_r_sa, h = h)
  alpha <- pmax(1e-10, forecast$model$par["alpha"])
  forecast$mean <- forecast$mean + (lsfit(0:(n - 1), x)$coefficients[2] / 2) * (0:(h - 1) + (1 - (1 - al
pha)^n) / alpha)

  # Reseasonalize
  if (seasonality) {
    forecast$mean <- forecast$mean * rep(tail(decomp$seasonality, frequency(x)), trunc(1 + h / fr
equency(x)))[1:h]
    forecast$fitted <- forecast$fitted * decomp$seasonality
  }
  forecast$residuals <- originalx - forecast$fitted
  forecast$method <- "Theta"

  return(forecast)
}
```

#Forecast model - predictions

```
#Return rate %
rrate_forecast <- thetamodel(datas$return_r_sa)
print(rrate_forecast) #predictions' values

##     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
##  95       9.952632 9.294699 10.51167 8.972586 10.83379
##  96       9.984981 9.176051 10.63032 8.791129 11.01524
##  97      10.017329 9.074215 10.73216 8.635384 11.17099
##  98      10.049677 8.983587 10.82278 8.496781 11.30959
##  99      10.082026 8.901123 10.90525 8.370663 11.43571
## 100      10.114374 8.824947 10.98142 8.254162 11.55221
## 101      10.146723 8.753809 11.05256 8.145365 11.66101
## 102      10.179071 8.686824 11.11955 8.042921 11.76345
## 103      10.211419 8.623340 11.18303 7.945831 11.86054
## 104      10.243768 8.562860 11.24351 7.853335 11.95304
## 105      10.276116 8.504994 11.30138 7.764835 12.04154
## 106      10.308465 8.449429 11.35694 7.679856 12.12652

plot(rrate_forecast) #plot
```




```
#company losses K euros
cost_forecast <- thetamodel(datas$cost_ret_sa)
print(cost_forecast)

##     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
##  95       9.930230 9.294699 10.51167 8.972586 10.83379
##  96       9.947923 9.176051 10.63032 8.791129 11.01524
##  97       9.965616 9.074215 10.73216 8.635384 11.17099
##  98       9.983309 8.983587 10.82278 8.496781 11.30959
##  99      10.001002 8.901123 10.90525 8.370663 11.43571
## 100      10.018695 8.824947 10.98142 8.254162 11.55221
## 101      10.036388 8.753809 11.05256 8.145365 11.66101
## 102      10.054081 8.686824 11.11955 8.042921 11.76345
## 103      10.071774 8.623340 11.18303 7.945831 11.86054
## 104      10.089467 8.562860 11.24351 7.853335 11.95304
## 105      10.107160 8.504994 11.30138 7.764835 12.04154
## 106      10.124853 8.449429 11.35694 7.679856 12.12652

plot(cost_forecast)
```




```
#profit K euros
profit_forecast <- thetamodel(datas$profit_sa)
print(profit_forecast)

##     Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
##  95       9.767097 9.294699 10.51167 8.972586 10.83379
##  96       9.678066 9.176051 10.63032 8.791129 11.01524
##  97       9.589034 9.074215 10.73216 8.635384 11.17099
##  98       9.500003 8.983587 10.82278 8.496781 11.30959
##  99       9.410971 8.901123 10.90525 8.370663 11.43571
## 100       9.321940 8.824947 10.98142 8.254162 11.55221
## 101       9.232909 8.753809 11.05256 8.145365 11.66101
## 102       9.143877 8.686824 11.11955 8.042921 11.76345
## 103       9.054846 8.623340 11.18303 7.945831 11.86054
## 104       8.965815 8.562860 11.24351 7.853335 11.95304
## 105       8.876783 8.504994 11.30138 7.764835 12.04154
## 106       8.787752 8.449429 11.35694 7.679856 12.12652

plot(profit_forecast)
```