Improve Coolblue's direct demand estimation model for substitutable products

Thesis - MSc Applied Mathematics

Josephine Clercx (4568095)

Improve Coolblue's direct demand estimation model for substitutable products

by

Josephine Clercx

to obtain the degree Master of Science in Applied Mathematics (specialisation: Statistics)
at the faculty of Electrical Engineering, Mathematics & Computer Science
at Delft University of Technology,
to be defended publicly on Monday August 22, 2022 at 10:30 AM.

Student number:	4568095	
Project duration:	December 13, 2021 – August 22, 2022	
Thesis committee:	Dr. A. J. (Annoesjka) Cabo,	TU Delft, Associate Professor
	Dr. A. F. F. (Alexis) Derumigny,	TU Delft, Assistant Professor
	Dr. ir. G. F. (Tina) Nane,	TU Delft, Assistant Professor

Abstract

This thesis aims to improve Coolblue's direct demand estimation model for substitutable products. Their current model consists of three sub-models which all provide their direct demand estimations. For every product, the direct demand is taken from one of the sub-models based on their performance in estimating the sales. The sub-models are the mean, linear and expectation-maximisation (EM) model. The linear model gives the most accurate expectations, whereas the mean model scores the lowest. Therefore, we have improved the mean model's estimations by creating a new estimator. Furthermore, we have investigated if an out-of-stock (OOS) period influences the sales and, therefore, the direct demand estimations. From this investigation, we conclude that for a part of the products, the sales are affected by OOS periods. However, these results are dependent on how they are investigated. Moreover, the OOS periods' influences are as likely to be positive as negative on the sales. Therefore it is challenging to react to these influences.

Preface

This thesis is my final work as a Delft University of Technology student. I started in September 2016 with the bachelor ”Technische Wiskunde” at EWI. This choice was easily made since I enjoyed mathematics in high school. After four years, from which one year consisted of a board year at my student association, I finished the bachelor in July 2020. In September 2020, I started with the master Applied Mathematics. My choice of specialisation was quickly made since I knew that I wanted to continue in the field of data and analytics. With this interest, I searched for a thesis project and contacted the Data Science department at Coolblue. And so on, I wrote my thesis there.

I would like to thank Alexis Derumigny for guiding me through this project. Next to supervising me, he taught me how to critically evaluate my own statistical work and improved my level of scientific writing. Also, special thanks to Tina Nane and Annoesjka Cabo for joining my thesis committee.

Furthermore, this thesis was impossible without the teamwork of the Supply Planning & Restock Management Data Science team of Coolblue. They gave me a wonderful introduction to the field of data science. I felt very welcomed by the team; they were always available to give feedback on my ideas. Especially, I want to thank Jaimy van Dijk with whom I had daily meetings to discuss my thoughts. Also, big thanks to Martin Skogholt and Giel van de Wiel, whom I met weekly to critically evaluate this thesis’s progress.

Josephine Clercx
Rotterdam, July 2022

Contents

1 Introduction 1
1.1 Coolblue . . . 1
1.2 Direct demand . . . 1
1.3 Coolblue's direct demand estimation model. . . . 2
1.4 Research questions . . . 2
1.5 Outline . . . 2

2 Data of the direct demand model 3
2.1 Structure of the data . . . 3
2.2 Sales analysis . . . 5
2.3 Other numerical variables . . . 5
2.4 Categorical variables . . . 5

3 The direct demand estimation model 7
3.1 Existing model . . . 7
3.1.1 Brief introduction . . . 7
3.1.2 The sub-models . . . 8
3.1.3 Model selection . . . 8
3.1.4 Post-processing steps . . . 8
3.2 Sub-model evaluation . . . 8
3.2.1 Correct splitting . . . 8
3.2.2 Applicability . . . 9
3.2.3 Best-performing sub-model . . . 9
3.2.4 Large errors in sales estimations . . . 10
3.2.5 Summary . . . 10
3.3 Improvements mean model . . . 11
3.3.1 New estimators . . . 11
3.3.2 Performance and comparison . . . 12

4 Out-of-stock period investigation 15
4.1 Motivation . . . 15
4.2 Additional notation . . . 15
4.3 Assumptions to investigate OOS periods . . . 16
4.4 Time-independent perspectives . . . 17
4.4.1 Perspectives for products with one OOS period . . . 17
4.4.2 Perspectives for products with multiple OOS periods. . . . 20
4.5 Time-dependent perspectives . . . 23
4.5.1 Dependence of the sales data . . . 23
4.5.2 Order of the $ARMA(p, q)$ processes . . . 24
4.5.3 Order of the $INGARCH(p, q)$ processes . . . 26
4.5.4 Perspectives for products with one OOS period . . . 27
4.5.5 Perspectives for products with multiple OOS periods. . . . 30
4.6 Perspectives for multiple products. . . . 30
4.7 Results . . . 31
4.7.1 Normalisation of the measures of interest. . . . 32
4.7.2 Results of the perspectives . . . 32
4.8 Conclusion . . . 39

5 Conclusion & Discussion 41
5.1 Discussion and further research . . . 42

Bibliography **43**

A Unpaired two-samples Wilcoxon test results **45**
A.1 Tests for Chapter 2 45
A.2 Tests for Chapter 3 46

B Detailed explanation of the EM model **47**

C Performance of the new estimators for the mean model per cluster **49**

D Derivations of equations **51**
D.1 Test statistic TS_{1,T^a} for comparing two Poisson distributions 51
D.2 Distance measure for $ARMA(1,1)$ 51
D.3 Test statistic TS_{4,T^a} for comparing two $ARMA(1,1)$ processes 52
D.4 Test statistic TS_{5,T^a} for comparing two $INGARCH(1,1)$ processes 53

E Additional figures and tables **57**
E.1 Overview of failures of the **arima()** function 57
E.2 Summaries of all the perspectives 57
E.3 Figures of the β_i^j investigation per product team 59

F Time-dependent perspectives for products with multiple OOS periods **61**

List of Figures

2.1 Example of the composition and levels of the product type Laptops. Note that there are more sub-product types and product segments than visualised in the figure. 4

3.1 An overview of the direct demand estimation model with $\boldsymbol{x}_{ij}$ the one-year historical product information vector, $\hat{\boldsymbol{s}}_{ij}^{Mean}$,$\hat{\boldsymbol{s}}_{ij}^{Linear}$,$\hat{\boldsymbol{s}}_{ij}^{EM}$ the vector of 90 days estimated sales of every sub-model and $\hat{\boldsymbol{d}}_{ij}^{Mean}$,$\hat{\boldsymbol{d}}_{ij}^{Linear}$,$\hat{\boldsymbol{d}}_{ij}^{EM}$ the vector of 90 days estimated direct demand of every sub-model and $\hat{\boldsymbol{d}}_{ij}^{*}$ is the estimated demand selected from one sub-model. 7
3.2 ECDF plot of the RRSEs per sub-model. These are the RRSEs of the estimated sales of the best performing sub-models per product. 11
3.3 ECDF of the RRSEs of all the estimations per estimator. The dotted lines are the average RRSEs. 13

4.1 An example of a product with three OOS periods. The blue and grey dotted lines are t_{jb} and t_{ja}, the last day before and the first day after OOS period j, respectively. 16
4.2 Simulated time series to show the difference in Perspectives 4.4.1, 4.4.2 and 4.4.3. The blue and grey dotted lines are t_{jb} and t_{ja}, the last day before and the first day after OOS period j, respectively. 19
4.3 Visualisation of the different alternative hypotheses for Perspective 4.4.6. The blue and grey dotted lines are t_{jb} and t_{ja}, the last day before and the first day after OOS period j, respectively.. 21
4.4 Results of the time-dependence investigation. 24
4.5 Heatmaps of $\upsilon_{AIC,p,q}$ and $\upsilon_{BIC,p,q}$ for $p, q = 0, ..., 15$, the percentages per $ARMA(p, q)$ fit that is selected as the best fit per product according to the AIC and BIC. 25
4.6 An example of time series with a fitted $ARMA(1, 1)$ process and its residuals. On the 20th of February there is an outlier, the residual on this date is marked with a square. . 28
4.7 Overview of the variables of Perspective 4.6.1. $a \rightarrow b$ means that a is dependent on b. . 31
4.8 Densities of the absolute normalised measures by the two normalisation methods introduced in Example 4.7.1. Note that because of the log scales the zero values of the measures are left out of the plot. 33
4.9 A Venn diagram of how many products reject time-independent perspectives. The products used in this Venn diagram are the products that could be tested for Perspectives 4.4.1, 4.4.2, 4.4.3 and 4.4.4. Note that we have combined the results of the perspective about larger or smaller in probability, since a product would not reject both these hypotheses. 35
4.10 A Venn diagram of how many products reject time-dependent perspectives. The products used in this Venn diagram are the products that could be tested for Perspectives 4.5.1.1, 4.5.1.2, 4.5.2, 4.5.3.1, 4.5.3.2, 4.5.4 and 4.5.5. 35
4.11 Box plots of the normalised measures of interest. The $ARMA(1, 1)$ distance metric measure and $INGARCH(1, 1)$ intervention effect measure only consist of positive values since the distance metric is squared and the intervention effect only can be positive by definition. 37
4.12 Scatter plot of the normalised means of the data before versus after an OOS period with identity line. 38
4.13 Density plots of the normalised non-zero means of the data before or after an OOS period . 38
4.14 Two dimensional kernel density estimations of the non-zero values of $\frac{\beta_i^j}{\sigma}$ and $\frac{\beta_i^{j+1}}{\sigma}$ with the actuals plotted. 38

4.15 Two scatter plots of the non-zero values of $\frac{\beta_i^j}{\sigma}$ versus $\frac{\beta_i^{j+1}}{\sigma}$, the first plot are the values from OOS periods that accept Perspective 4.4.3, the second plot are values that reject this perspective. 38
4.16 Two dimensional kernel density estimation of $\frac{\beta_i^j}{\sigma} - \frac{\beta_i^{j+1}}{\sigma}$ and the middle day of the OOS period (= med($\mathcal{T}_i^{*j}$)). 39

C.1 The silhouette index for every cluster size. 49
C.2 The average daily sales of the four clusters plotted as time series. 50

E.1 Two dimensional kernel density estimations of the non-zero values of $\frac{\beta_i^j}{\sigma}$ and $\frac{\beta_i^{j+1}}{\sigma}$ for every product team. 60
E.2 P-values as an heat map of the test if the correlations differ significantly. Some on the product team names are abbreviated. 60

List of Tables

2.1 Percentages of how many variables are known in the dataset. 5

3.1 Percentage of products for which the sub-models are applicable, for each country j. . . 9
3.2 Percentage of the products for which the sub-model has the most accurate estimations, based on the RMSE of the estimated sales. The four subsets contain 27.83%, 8.59%, 41.01% and 22.51% of the products, respectively. 10
3.3 Overview of the sub-model performance on the three criteria. ”+++” indicates scoring the best on this criteria, whereas ”+” indicates scoring the most minor. 10
3.4 Percentages of products where the estimation is replaced by zero due to the lack of data. 12

4.1 Results of the example product in Figure 4.2. 19
4.2 An overview of the measures of interest that will be investigated, with $i = 1, ..., n$ the product and $j = 1, ..., k_i + 1$ the OOS period. Together with the perspective from which the measure comes. 30
4.3 Percentage of the products with the number of OOS periods that meet the assumptions introduced in Section 4.3. 33
4.4 The results of the perspectives for products with one OOS period. Only Perspective 4.4.5 is omitted since these results are investigated separately. The ”Applicability percentage” is the percentage of products which the perspective could investigate. The ”Rejection percentage of applicable products” is the percentage of applicable products that reject the hypothesis in the perspective. The ”Rejection percentage of 5.01 % of the products” is another rejection percentage, but of the subset of products that all perspectives could investigate (except Perspective 4.4.5). These are 5.01 % of all products. 34
4.5 The results of the time-independent perspectives on the slow-moving products. These are 39.53 % of the products. The time-dependent perspectives are not considered since the used time series are not created especially for binary data. 36
4.6 The results of the perspectives for products with multiple OOS period. Only Perspective 4.4.10 is omitted. The ”Applicability percentage” is the percentage of products which the perspective could investigate. The ”Rejection percentage of applicable products” is the percentage of applicable products that reject the alternative hypothesis in the perspective, per alternative hypothesis. The ”Rejection percentage of 3.39 % of the products” is another rejection percentage, but of the subset of products that all perspectives could investigate (except Perspective 4.4.10). These are 3.39 % of all products. Some of the results of hypothesis $H_{1.b}$ miss because this hypothesis is not tested. 36
4.7 The results of Perspective 4.6.1 for every measure of interest. The ”Applicability percentage” is the percentage of all the values of the measure which were available to use. 37
4.8 Overview of how many percentage of the normalised means equal zero or are larger than zero. 38
4.9 Correlation between $\frac{\beta_i^j}{\sigma}$ and $\frac{\beta_i^{j+1}}{\sigma}$ for every product team. 39

A.1 Results of the significance tests 1 till 4. 46
A.2 Results of the significance test 5. 46
A.3 Results of the significance tests 6 till 8. 46

E.1 Percentage of the products per p and q value for which the function **arima()** failed to fit an $ARMA(p, q)$ process due to different reasons. 57
E.2 A summary of all the time-independent perspectives: A short description of the perspective with the tests per alternative hypothesis. 58

E.3 A summary of all the time-dependent perspectives: A short description of the perspective with the tests per alternative hypothesis. 59

1 Introduction

In this thesis, we will investigate Coolblue's direct demand estimation model for substitutable products and improve this model in multiple aspects. First, we will briefly introduce the company Coolblue. Then, we will explain the concept of direct demand and why this is essential for Coolblue. After that, Coolblue's direct demand estimation model is described shortly. Finally, the research questions are presented, followed by an outline of the thesis.

1.1. Coolblue

Coolblue is a Dutch e-commerce company founded in 1999. They are one of the largest online retailers in The Netherlands and Belgium today. In 2020 they expanded their market to Germany and have grown there since then. Their main office is in Rotterdam, and the central warehouse is in Tilburg. They also have 23 physical stores distributed over the three countries.

Many consumer electronics are delivered by Coolblue's delivery service consisting of vans and electric bikes. They are well-known for their excellent service. Additionally, they sell solar panels and charging stations, and offer Home Office Stores for employees of other companies. In 2021 Coolblue announced their own energy label called Coolblue Energy.

1.2. Direct demand

It is a big challenge for Coolblue to find a balance between having enough products available to serve the customer demand and the costs of keeping all the products stored in warehouses. To get to this optimum, we need to know how much of a product the customers want, this is called the direct demand. Moreover, the direct demand is also used to forecast the sales and review the past performance of Coolblue.

One of the challenges when measuring the direct demand is when the demand cannot be met, this results in a lost sale which is not observed. An example of a situation with lost sales: Assume that at the beginning of a period, there are 20 products available, and there will be no extra incoming deliveries in this period. If there are still eight products left in stock at the end of the period, we know that the demand was 12. However, if there are zero products left in stock at the end of the period, we only know that the demand was at least 20, but we do not know if we lost sales.

Another challenge when measuring direct demand is the substitution of products. If a product is out of stock, the customer could go for another product that is almost similar (another colour, version, brand etc.). As a result, the sales of a product in a certain period are not a good representation of the direct demand because it consists of the direct demand and the substitution demand.

To conclude, a product's direct demand equals the sales when the product and all the product's substitutes are in stock.

1.3. Coolblue's direct demand estimation model

Coolblue's direct demand estimation model estimates every day the direct demand for the last 90 days for every product per country. The model consists of three sub-models: the mean model, the linear model and the expectation-maximisation (EM) model. Every sub-model estimates the direct demand and sales, if possible. Then, the estimated direct demand per product per country is taken from one sub-model. This choice is based on the performance of the sub-models in estimating the sales since this variable is observed. Finally, every product per country has its estimated direct demand for the last 90 days coming from one of the sub-models.

1.4. Research questions

To investigate and improve the direct demand estimation model, we begin by analysing the data used in the model. After that, we compare the performance of the three sub-models (mean mode, linear mean and EM model). From this comparison, we conclude that the mean model can be improved. Moreover, experts on Coolblue's direct demand estimation model advise investigating if an out-of-stock (OOS) period influences the sales afterwards because they expect an influence. Using this information, we try to answer the following research questions in this study:

- *How do the three sub-models perform compared to each other?*
- *Can we improve the estimated demand provided by the mean model?*
- *Do out-of-stock periods influence the sales? And if so, how can we react to this influence such that it does not affect the direct demand estimations.*

1.5. Outline

In the following chapters, we will answer the research questions introduced above. First, a thorough description and analysis of the data input of the direct demand model are given in Chapter 2 to introduce the reader further to Coolblue. After that, the direct demand model is explained in detail in Chapter 3, followed by a sub-model evaluation to answer the first research question. Additionally, we answer the second research question at the end of this chapter. Finally, in Chapter 4 the last research question is answered with an extensive out-of-stock investigation. In the final chapter, we conclude and discuss our research and give recommendations for further research.

All the simulations, figures, and results are made in the programming language R. The codes are not public because of confidence regulations, they can be requested by mailing to `josephineclercx@hotmail.com`.

2

Data of the direct demand model

This chapter describes the data used in Coolblue's direct demand estimation model. First, we introduce all the variables of the dataset. After that, we share analysis of the sales variable since this is an essential variable in the model. Finally, we analyse the other variables briefly.

2.1. Structure of the data

For every product $i = 1, ..., n$ in country $j = \{NL, BE, GER\}$ on day $t = t_1, ..., t_{T_{ij}}$ product information is used for the direct demand model. This is bundled in vector $\boldsymbol{x}_{ijt}$. If known, 365 days of historical data is given as input, else the subset of recent historical data is given. Below are the variables of the vector listed. We only provide the mathematical notation of the variables if we use them in the thesis. Moreover, the data used for investigation and results in this thesis is the product information from 12/21/2020 until 12/20/2021.

- Sales ($s_{ijt} \geq 0$): Some sales are removed: the cancelled orders, non-standard products (like insurances, services, etc.) and bundled products. The sales do include pre-ordered sales.
- Status (st_{ijt}): To give products an availability status, the statuses of the products on Coolblue's website are used. Products can have status available, unavailable, permanently unavailable or pre-order. See Definition 2.1.1 for the concrete definition of available and unavailable. The definitions of permanently unavailable and pre-order are not concrete.

  Definition 2.1.1: Available and unavailable status

  – A product has status *available* on day t if it is in stock on the website for the whole day.
  – A product has status *unavailable* on day t if it is out of stock for at least one moment on the whole day.

  $st_{ijt} \in ST = \{\text{available, unavailable, permanently unavailable, pre-order}\} = \{A, U, Perm.\text{-}U, P\text{-}O\}$.

- Status-2 (st^*_{ijt}): This is another status that Coolblue uses, we name it status-2. The status-2 of products can be partial available or partial unavailable. See Definition 2.1.2 for the concrete definition of these statuses-2. The reason that there is another status definition is because Definition 2.1.1 results in product being sold when the product has status unavailable, which is not logical.

  Definition 2.1.2: Partial available and partial unavailable status-2

  – A product has status-2 *partial available* on day t if it has status available or if there is at least one sale on day t.
  – A product has status-2 *partial unavailable* on day t if it has status unavailable and no sales on day t, or when it has status permanently unavailable or pre-order.

$st^*_{ijt} \in ST^* = \{$partial available, partial unavailable$\} = \{PA, PU\}$.

- Segment composition: Describes the composition of each product type, sub-product type and product segment. It specifies which product belongs in which segment, for an example, see Figure 2.1.
- Market share ($m_{ijt} \geq 0$): Via an external source, Gfk [1], the market share of a product segment is known. The market share represents how much percentage Coolblue contributed to the total market. It is calculated per week per product segment. However, there are certain caveats; it is not available for all product segments. Also, it can be unreliable; the data Coolblue sends to Gfk is not always the same as what Gfk reports back.
- Website visits ($v_{ijt} \geq 0$): These are direct page landings on the website.
- Price ($p_{ijt} > 0$): The price of a product in euros.
- Discount price: If there is a discount or a price change on a product, the discount price is the product's price before the discount, or the price change happens.
- Competitor price: The prices of products of the competitors if they are known.
- Parent product sales ($pt_{ijt} \geq 0$): A product's parent product is another product linked to the product as a product likely to be bought with it. The parent product is the "main" product, the product itself is a smaller product that can be useful by the parent product. An example: The parent product of a phone case is the phone.
- Dutch national holidays: This is an indicator variable which indicates if there is a Dutch national holiday.
- Coolblue's choice (cc_{ijt}): This is a label of Coolblue which indicates that a product is the best product compared to similar products, according to experts of Coolblue. A product can be Coolblue's choice all the time, for a limited time or never.
  $cc_{ijt} \in \{TRUE, FALSE\}$

Figure 2.1: Example of the composition and levels of the product type Laptops. Note that there are more sub-product types and product segments than visualised in the figure.

Not all the variables listed above are known for every product or date. Table 2.1 is an overview of which variables are not always known per country j. The reason why competitors' prices are unknown sometimes is that not every product is being sold at a competitor, or Coolblue does not have the capability to know the prices of competitors for every product. Moreover, the discount price is sometimes unknown because there is no discount price for a product. The product itself has only its price. Lastly, the market share is unknown sometimes because Gfk does not provide the market share for every

[1] GfK is a provider of data and analytics to the consumer goods industry, www.gfk.com

product. Comparing the percentage of known variables per country shows that Germany has the least known variables. The reason for this at the competitor prices and market share variables is that Coolblue had entered the market of Germany recently. Therefore, Coolblue's market in Germany is not as developed as in the Netherlands and Belgium. As a result, this information is not known yet.

Percentage of how many variables are known			
Variable	**NL**	**BE**	**GER**
Competitor price 1	51.85	49.28	42.69
Competitor price 2	35.18	31.40	30.99
Discount price	15.05	14.22	12.25
Market share	32.57	32.00	9.63

Table 2.1: Percentages of how many variables are known in the dataset.

2.2. Sales analysis

Author's note: This section is confidential.

2.3. Other numerical variables

Author's note: This section is confidential.

2.4. Categorical variables

Author's note: This section is confidential.

3

The direct demand estimation model

This chapter describes and improves Coolblue's direct demand model. First, in Section 3.1 is the existing direct demand model of Coolblue shared in detail. Then, in Section 3.2 we evaluate the three sub-models of the direct demand model. Finally, in Section 3.3 we improve the estimations of the mean model.

3.1. Existing model

The model of Coolblue to estimate the direct demand for products is described in this section. A brief introduction of the model will be given first. Then, we describe how the sub-models estimate the direct demand. After that, the processing stages of the model are explained.

3.1.1. Brief introduction

Coolblue currently has a model to daily estimate direct demand for every product in each country for the past 90 days. The input data for every product i in country j is the product information vector $\boldsymbol{x}_{ijt}$ of the recent 365 days. The model is a combination of three sub-models: the mean model, the linear model and the expectation maximisation (EM) model. All of the sub-models estimate the direct demand and the sales for the products. To choose the most accurate demand estimation, the sales estimations are evaluated. This is done because the sales are observed while the direct demand is not. The model with the most accurate estimated sales for a product is the one from which the estimated direct demand is taken. After that, two post-processing are applied to scale the estimates if needed. As a result, every product has its estimated direct demand from one of the three sub-models. Figure 3.1 is an overview of the direct demand estimation model.

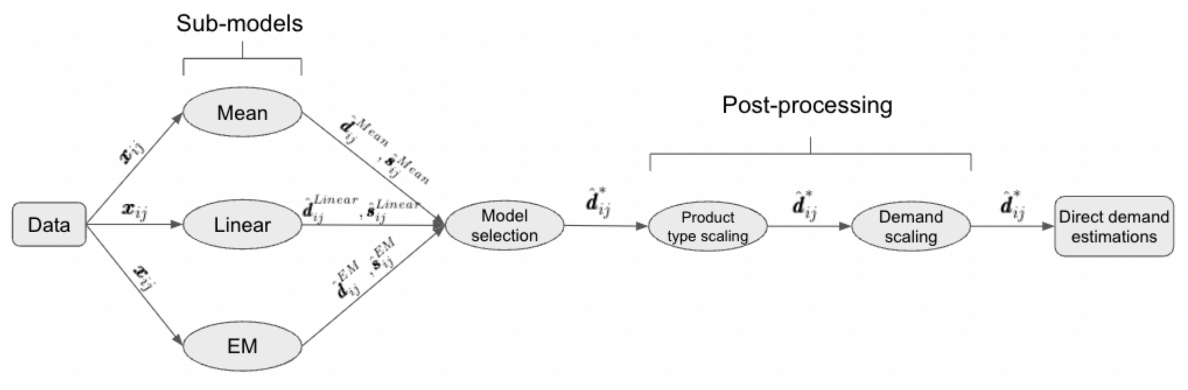

Figure 3.1: An overview of the direct demand estimation model with $\boldsymbol{x}_{ij}$ the one-year historical product information vector, $\hat{\boldsymbol{s}}_{ij}^{Mean}, \hat{\boldsymbol{s}}_{ij}^{Linear}, \hat{\boldsymbol{s}}_{ij}^{EM}$ the vector of 90 days estimated sales of every sub-model and $\hat{\boldsymbol{d}}_{ij}^{Mean}, \hat{\boldsymbol{d}}_{ij}^{Linear}, \hat{\boldsymbol{d}}_{ij}^{EM}$ the vector of 90 days estimated direct demand of every sub-model and $\hat{\boldsymbol{d}}_{ij}^{*}$ is the estimated demand selected from one sub-model.

Coolblue uses BigQuery as data warehouse to store the input and output data of their models, the direct demand model is ran in Python.

3.1.2. The sub-models

Author's note: This section is confidential.

3.1.3. Model selection

Author's note: This section is confidential.

3.1.4. Post-processing steps

Author's note: This section is confidential.

3.2. Sub-model evaluation

To improve Coolblue's direct demand estimation model, we will look for weaknesses in the sub-models' performance. This investigation is covered in the first research question that we will answer in this section:

- *How do the three sub-models perform compared to each other?*

They are compared on three criteria: applicability, best-estimating and occurrences of large errors. The sales estimations of the sub-models are evaluated for this. These estimations come together in the model selection step. When we investigated the model selection step, we detected incorrectness in splitting the train and test set. Therefore, we will first share the splitting process and advise on a better process. After that, we will go through the three criteria points.

3.2.1. Correct splitting

Every sub-model's sales estimations are made with a train set and evaluated on a test set in the model selection step. The test set should only include days with status available since Coolblue wants to evaluate the estimations only on these days. When analysing this splitting process, we detected that the train set is not representable: It does not consist of the same data as the data set used to estimate demand. Algorithm 1 explains the current incorrect splitting method for every product's data.

Algorithm 1 The incorrect splitting of the train and test set in the model selection step

Initialisation

$\mathcal{T}_{ij} = \{t_1, ..., t_{T_{ij}}\}$ $\quad i = 1, ..., n \quad j \in C = \{NL, BE, GER\}$

$\mathcal{T}_{ij}^{\text{available}} = \{t | st_{ijt} = \text{available}\} \subset \mathcal{T}_{ij}$

Splitting

$\mathcal{T}_{ij}^{\text{proposed test set}} = \{\text{sample 20\% uniformly and random without replacement of } \mathcal{T}_{ij}\}$

$\mathcal{T}_{ij}^{\text{test set}} = \mathcal{T}_{ij}^{\text{proposed test set}} \cap \mathcal{T}_{ij}^{\text{available}}$

$\mathcal{T}_{\text{train set}}^{ij} = \mathcal{T}_{ij}^{\text{available}} \setminus \mathcal{T}_{ij}^{\text{test set}}$

The train set consists only of days that have status available, this influences the estimations of the mean and EM model: They both should use all the historical data, regardless of the status. As a result, these sub-models' estimation methods do not align with the demand estimation method. Moreover, how the test set is constructed results in a higher chance of having an empty test set, this can be improved. Therefore, we advise an improved splitting method, see Algorithm 2.

With this splitting method, the train set consists of random days regardless of the status. Also, the test set has a smaller chance of being an empty set, however if there are no days with status available or status-2 partially available, the test set will not consist of days with these statuses. This is not entirely how Coolblue wants to evaluate the estimations, they can change this if wanted.

In the following sections, we will compare the sub-models with the criteria introduced. The output data of the model selection step is used where the old splitting method (Algorithm 1) still is applied, with

Algorithm 2 An advice to improve the splitting of the train and test set in the model selection step

Initialisation

$\mathcal{T}^{ij} = \{t_1, ..., t_{T_{ij}}\}$ $\quad i = 1, ..., n \quad j \in C = \{NL, BE, GER\}$

$\mathcal{T}_{ij}^{\text{available}} = \{t | st_{ijt} = \text{available}\} \subset \mathcal{T}_{ij}$

$\mathcal{T}_{ij}^{\text{partial available}} = \mathcal{T}_{ij}^{\text{available}} \cup \{t | s_{ijt} > 0\}$

Splitting

if $|\mathcal{T}_{ij}| = 1$ **then**

product i for country j is not being evaluated

end if

if $\mathcal{T}_{ij}^{\text{available}} \neq \emptyset$ **then**

$\mathcal{T}_{ij}^{\text{test set}} = \{$sample 20% (minimal one element) of $\mathcal{T}_{ij}^{\text{available}}$ uniformly and at random without replacement$\}$

else if $\mathcal{T}_{ij}^{\text{partial available}} \neq \emptyset$ **then**

$\mathcal{T}_{ij}^{\text{test set}} = \{$sample 20% (minimal one element) of $\mathcal{T}_{ij}^{\text{partial available}}$ uniformly and at random without replacement$\}$

else

$\mathcal{T}_{ij}^{\text{test set}} = \{$sample 20% (minimal one element) of $\mathcal{T}_{ij}$ uniformly and at random without replacement$\}$

end if

$\mathcal{T}_{ij}^{\text{train set}} = \mathcal{T}_{ij} \setminus \mathcal{T}_{ij}^{\text{test set}}$

the same input data as in the previous chapter: Product information from 12/21/2020 until 12/20/2021.

3.2.2. Applicability

Not every sub-model can make an estimation for every product because of the lack of data, as told before. Table 3.1 consists of the percentages of products for which the model is applicable per country. All products can be estimated by the mean model, whereas the linear and EM model cannot. Especially the EM model can estimate only a minority of the products. Moreover, German product data is the least known compared to the Netherlands and Belgium because of Coolblue's recent entry into the German market.

Percentage of the products for which the sub-models are applicable			
Country \ Sub-model	**Mean model**	**Linear model**	**EM model**
NL	100.00	68.90	36.41
BE	100.00	64.00	35.28
GER	100.00	28.92	16.95

Table 3.1: Percentage of products for which the sub-models are applicable, for each country j.

The results of the data analysis in Chapter 2 and Table 3.1 are nearly the same for the Netherlands and Belgium. The German results are mostly slightly different because of the recent market entry. Therefore, we choose to use only the data of products in the Netherlands for the following sections and chapters to reduce the computation time of generating results. The Dutch products are a good representation of all the products.

3.2.3. Best-performing sub-model

The second criteria which we use to investigate the sub-models is their performance in estimating the sales, the root-mean-squared error (RMSE) is used as metric. Per set of sub-models that can estimate the sales for a product, the percentage of products for which the sub-model is the most accurate in estimating is given in Table 3.2. The linear model gives the most accurate estimations overall, whereas the mean and EM model perform the same approximately.

Percentage of products for which the model is selected as best-performing sub-model			
Demand estimated by	**Mean model**	**Linear model**	**Mean model**
All three sub-models	21.57	57.66	20.77
Mean & EM model	49.62	-	50.38
Mean & Linear model	39.99	60.01	-
Mean model	100.00	-	-

Table 3.2: Percentage of the products for which the sub-model has the most accurate estimations, based on the RMSE of the estimated sales. The four subsets contain 27.83%, 8.59%, 41.01% and 22.51% of the products, respectively.

3.2.4. Large errors in sales estimations

Finally, we investigate the occurrences of larger errors in the estimations to detect which sub-model makes poor estimations. Only the estimations of the best-performing sub-model per products are used, thus only one estimation per product. To detect larger errors, we use another metric than the RMSE, we use the root-relative-squared error (RRSE) [21]. It is defined by:

$$RRSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\max\left(\sum_{i=1}^{n}(y_i - \bar{y})^2, \epsilon\right)}}, \tag{3.1}$$

with $\hat{y}_i$ the prediction of observation y_i, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, the mean of the n observations, and ϵ is a small positive number to avoid division by zero, we choose $\epsilon = 1 * 10^{-4}$. We have multiple reasons to use this metric: This metric is relative, it does not depend on the size of the product's sales; it weights large errors more heavenly; and the chance that the value $\sum_{i=1}^{n}(y_i - \bar{y})^2$ in the denominator equals zero is smaller than other metrics. This last reason is important since the sales equal zero frequently, as seen before in Chapter 2.

Figure 3.2 plots the empirical cumulative distribution function (ECDF) of the RRSEs per sub-model. If we look at the highest 25 % values of the RRSE per sub-model, the mean model has the largest values of RRSE, followed by the EM and linear model. Moreover, the highest 25 % RRSEs of the mean model are significantly higher than those of the EM and linear model, and the highest 25 % RRSEs of the EM model are significantly higher than those of the linear model. See Chapter A for the details (tests 6-8). We can conclude that the linear model also has the best results in this criteria.

3.2.5. Summary

Table 3.3 gives an overview of the performance of the three sub-models on our criteria. The linear model has the best score on average, it could only be improved on applicability. On the contrary, the mean model can estimate all products, but it scores poorly on estimating the sales. The EM model scores average: Its estimations are medium, but the sub-model cannot estimate many products.

Summary of the sub-model evaluation			
Sub-model	**Mean model**	**Linear model**	**EM model**
Applicibility	+++	++	+
Best-performing	+	+++	+
Large errors	+	+++	++

Table 3.3: Overview of the sub-model performance on the three criteria. "+++" indicates scoring the best on this criteria, whereas "+" indicates scoring the most minor.

The result that the linear model outperforms the other two sub-models while using the most data of all sub-models is interesting. We can state that in these sub-models, the amount of data increases the accuracy of the estimations. This is in line with other topics in statistics. For example, the law of large numbers states that when the number of observations increases, the sample mean approaches the theoretical mean [19].

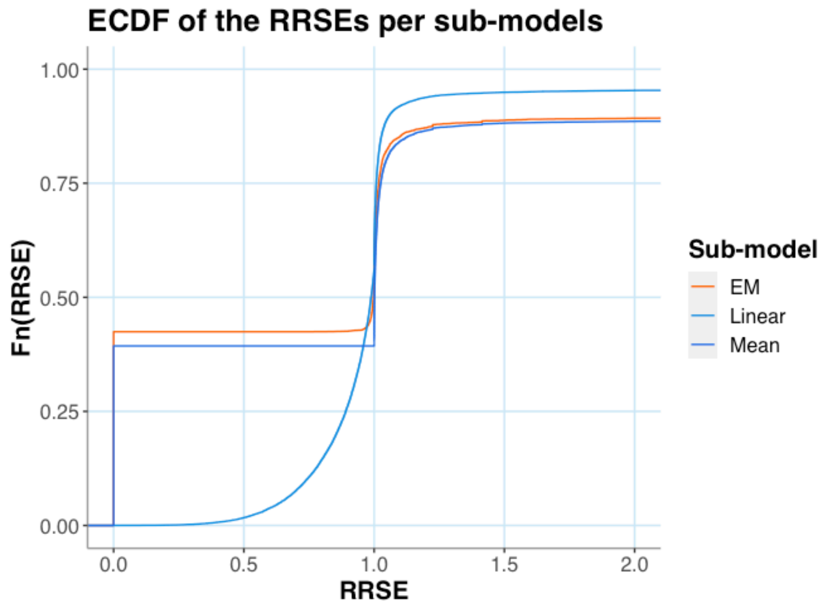

Figure 3.2: ECDF plot of the RRSEs per sub-model. These are the RRSEs of the estimated sales of the best performing sub-models per product.

3.3. Improvements mean model

In the previous section, we concluded that the mean model scores poorly on estimating, while it is a sub-model that is the backup sub-model if the other two sub-models fail to make estimations due to the lack of external data. Therefore, we want to improve the estimations made by the mean model. We will answer the second research question of this thesis in this section:

- *Can we improve the estimated demand provided by the mean model?*

We will improve the estimated demand by changing the estimator in Definition **??**. There are two criteria we want to meet when changing the estimator:

- Keep the same applicability as the current estimator: We want the mean model to be still applicable for all the products. To do this, we should not increase the needed data since the likelihood of failure increases when external data is included in the model. Therefore, only the sales and status of the products should be used.
- Reduce the average RRSE of the sales estimates of all the products in the Netherlands. The current model has an average RRSE of 14.00, we want to be below this number with the new estimator.

We have created three estimators that meet the first criteria point. We will investigate which new estimator has a lower average RRSE than 14.00. Note that in the current estimator, the formula for estimating the demand and sales are the same, this will also be in the new estimators. We will compare the performance of the estimators on their sales estimates, therefore we describe the new estimators as sales estimators

3.3.1. New estimators

The current estimator takes the mean of the sales on all days, even when a product has status unavailable. As a result, the mean can be taken over many days when a product cannot be sold because it is not in stock. This is our inspiration for the new estimators.

The first estimator equals the mean of the historical sales on the days when the product has status available. If a product never has status available, the estimator equals zero. The vector of 90 sales estimates for product i in country j can be written as Definition 3.3.1.

Definition 3.3.1: New estimator 1 mean model

$$\hat{\boldsymbol{s}}_{ij}^{\text{New estimator 1 mean model}} = \max\Big(\frac{1}{T_{ij}} \sum_{t=t_1}^{t_{T_{ij}}} s_{ijt} \mathbb{1}_{\{t \mid st_{ijt}=\text{available}\}}, 0\Big). \tag{3.2}$$

The second estimator equals the mean of the historical sales, with a weight on the days a product has status unavailable to compensate for the days when a product is not in stock. When a product has status pre-order or permanently unavailable, these days are not used. The vector of 90 sales estimates for product i in country j can be written as Definition 3.3.2.

Definition 3.3.2: New estimator 2 mean model

$$\hat{\boldsymbol{s}}_{ij}^{\text{New estimator 2 mean model}} = \max\Big(\frac{1}{T_{ij}} \sum_{t=t_1}^{t_{T_{ij}}} \big(s_{ijt} \mathbb{1}_{\{t \mid st_{ijt}=\text{available}\}} + 0.5 * s_{ijt} \mathbb{1}_{\{t \mid st_{ijt}=\text{unavailable}\}}\big), 0\Big). \tag{3.3}$$

The third estimator equals the mean of the historical sales when the product has status-2 partial available. If a product is never partial available, the estimator equals zero. The vector or 90 sales estimates for product i in country j can be written as Definition 3.3.3.

Definition 3.3.3: New estimator 3 mean model

$$\hat{\boldsymbol{s}}_{ij}^{\text{New estimator 3 mean model}} = \max\Big(\frac{1}{T_{ij}} \sum_{t=t_1}^{t_{T_{ij}}} s_{ijt} \mathbb{1}_{\{t \mid st^*_{ijt}=\text{partial available}\}}, 0\Big). \tag{3.4}$$

For the next section, we will name the current estimator as defined in Definition **??** as the ”Old estimator” and the new estimators introduced above as ”Estimator 1”, ”Estimator 2”, and ”Estimator 3”, in the same order as they are listed above. Moreover, the data used to compare the estimators is the product information from 12/21/2020 until 12/20/2021 of products sold in the Netherlands.

3.3.2. Performance and comparison

First, we are interested in how many times one of the estimators changed the estimation to zero because there are no days a product has status available/partial available. In Table 3.4 are the percentage of all products for which the estimation has been replaced by zero per estimator. The estimates of the Old estimator and Estimator 2 are never replaced since those estimators also use days when the product has status unavailable. On the contrary, 18.04 % and 15.63 % of the estimations of Estimators 1 and 3, respectively, are replaced by zero.

Percentage of the products where the estimate has been imputed	
Estimator	**Percentage**
Estimator 1	18.04
Estimator 2	0.00
Estimator 3	15.63
Old estimator	0.00

Table 3.4: Percentages of products where the estimation is replaced by zero due to the lack of data.

Furthermore, we want to know if the performance of the estimators is influenced by the type of products. We have clustered the products based on their sales and compared the estimations of the estimators. Out of these results, the same estimator has the best estimations. Therefore, the type of products does not influence the performance of these estimators, and we will not distinguish between the type of products in this investigation. The detailed investigation is in Chapter C.

Finally, we investigate the performance of estimators: We compare the RRSE of the estimated sales of every estimator. In Figure 3.3 the ECDFs of the RRSEs of every estimator, including the means, are plotted. The average RRSE of the Old estimator is the highest, thus we have accomplished our goal to decrease the average RRSE of 14.00 with all the new estimators. Moreover, the new estimators' outliers in RRSEs show that Estimator 2 has the least outliers. However, Estimator 2 does not have the least average RRSE, Estimator 1 has. Our goal is to decrease the average RRSE, therefore we advise replacing the Old estimator with Estimator 1, although it has more outliers than Estimator 2.

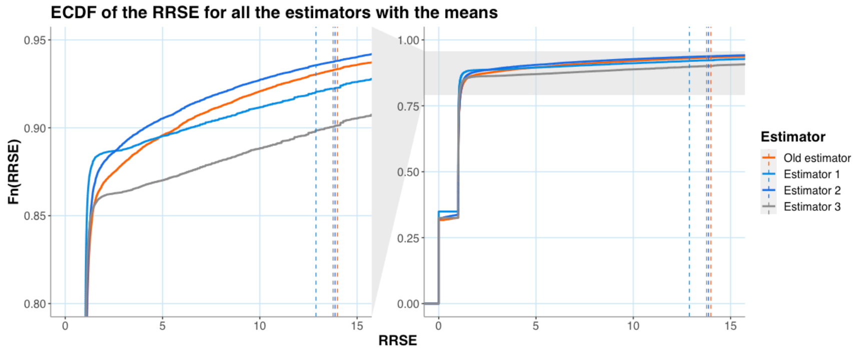

Figure 3.3: ECDF of the RRSEs of all the estimations per estimator. The dotted lines are the average RRSEs.

4

Out-of-stock period investigation

Coolblue's direct demand model experts advise investigating if a product being out-of-stock (OOS) for a period influences the sales after such a period since they expect an influence. An OOS period is a period when a product has status-2 partial unavailable for consecutive days. We will investigate this topic by answering the last research question of this thesis in this chapter:

- *Do out-of-stock periods influence the sales? And if so, how can we react to this influence such that it does not affect the direct demand estimations.*

First, we will elaborate more on our motivation to investigate this subject. Then, we introduce additional notation that is needed for this chapter. After that, we share our assumptions and investigation to answer this research question, and finally, we answer the question.

4.1. Motivation

Multiple scenarios occur when customers find out their product is not in stock. For example, they can delay their purchase and come back when the product is back in stock. As a result, the sales suddenly increase when the product is back in stock. However, it can also happen the other way around: they find a substitute of the same brand or from another brand, or they buy the product at another store or do not buy the product at all [6] [9] [10] [23]. These responses result in a decrease in sales after an OOS period. Therefore, we are interested in increases or decreases in sales after an OOS period.

These increases or decreases in sales can affect the demand estimates. Because first of all, sales are one of the variables on which these estimates are based on. If customers delay their purchase, the peak of sales after an OOS period can equal the summed-up demand of the OOS period. The direct demand model could interpret this wrong, resulting in inaccurately estimated demands. Furthermore, these sales peaks or valleys could distort the model selection process. In this process, the best-performing model is chosen for every product. The performing results could be distorted by unexpected peaks and valleys in the actual sales. This results in incorrectly chosen sub-models by the evaluation process, resulting in a less well-estimated demand.

4.2. Additional notation

In Chapter 2 we already introduced notation, we will elaborate on it for this chapter. Note that we will not distinguish between the products in different countries, therefore we will omit the subscript for the country in this chapter. For example, s_{ij} becomes s_i.

- Let $n \in \mathbb{N}^*$ be the number of products that we investigate.
- Let $T_i \in \mathbb{N}^*$ be the length of the dataset for product $i = 1, ..., n$.
- Let $k_i \in \mathbb{N}^*$ be the number of OOS periods that we investigate for product $i = 1, ..., n$.
- Let $\mathcal{T}_i$ be the set of days for which we have data of product $i = 1, ..., n$. Thus, $\mathcal{T}_i = \{t_1, ..., t_{T_i}\}$. This set can be partitioned into multiple sets: $\mathcal{T}_i = \{\mathcal{T}_i^1, \mathcal{T}_i^{*1}, \mathcal{T}_i^2, \mathcal{T}_i^{*2}, \mathcal{T}_i^3, ..., \mathcal{T}_i^{*k_i}, \mathcal{T}_i^{k_i+1}\}$, with $\mathcal{T}_i^j$ the set

of consecutive days a product has status-2 partial available for $j = 1, ..., k_i + 1$ and $\mathcal{T}_i^{*l}$ the set of consecutive days a product has status-2 partial unavailable for $l = 1, ..., k_i$.

- Let $\mathcal{T}_i^{\text{partial available}} = \{\mathcal{T}_i^1, \mathcal{T}_i^2, ..., \mathcal{T}_i^{k_i}, \mathcal{T}_i^{k_1+1}\}$ and $\mathcal{T}_i^{\text{partial unavailable}} = \{\mathcal{T}_i^{*1}, \mathcal{T}_i^{*2}, ..., \mathcal{T}_i^{*k_i}\}$ be the sets of in-stock and out-stock days, respectively.
- Let $T_i^j = |\mathcal{T}_i^j|$ and $T_i^{*j} = |\mathcal{T}_i^{*j}|$ be the lengths of the datasets.
- Let $t_{i,b}^j = \max\{\mathcal{T}_i^j\}$ and $t_{i,a}^j = \min\{\mathcal{T}_i^{j+1}\}$ be the day before and after OOS period $j = 1, ..., k_i$, respectively. For example, the set of days for product i with $k_i = 1$ can be described as:

$$\mathcal{T}_i = (\mathcal{T}_i^1, \mathcal{T}_i^{*1}, \mathcal{T}_i^2) = (t_1, ..., t_{i,b}^1, t_{i,b+1}^1, ..., t_{i,a-1}^1, t_{i,a}^1, ..., t_{T_i}).$$

Note that $s_{it} = 0$ whenever $t \in \mathcal{T}_i^{*j}$ for all j, since a product cannot be sold when it has status-2 partial unavailable. To make the notation more clear, we have created an example product with three OOS periods, see Figure 4.1.

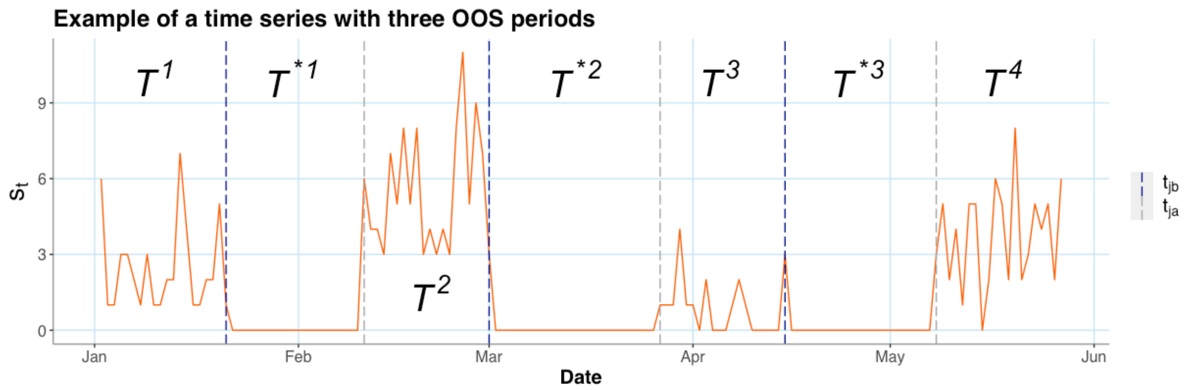

Figure 4.1: An example of a product with three OOS periods. The blue and grey dotted lines are t_{jb} and t_{ja}, the last day before and the first day after OOS period j, respectively.

We simulated $s_t = (X_1, ..., X_{20}, U_1, ..., U_{20}, Y_1, ..., Y_{20}, U_{21}, ..., U_{46}, Z_1, ..., Z_{20}, U_{47}, ..., U_{59}, W_1, ..., W_{20})$ with $X_i \sim Pois(3)$, $U_i = 0$, $Y_i \sim Pois(6)$, $Z_i \sim Pois(1)$ and $W_i \sim Pois(4)$.

4.3. Assumptions to investigate OOS periods

To investigate an OOS period, we have created assumptions about what an OOS period is. These assumptions are about the length of an OOS period and the length of the data before and after an OOS period. These assumptions are: We investigate OOS period j of product i if

- $T_i^{*j} \geq 14$, else $\mathcal{T}_i^{*j}$ is removed from the dataset,
- and if $T_i^j \geq 7$ and $T_i^{j+1} \geq 7$, else OOS period j is not investigated.

This choice of the minimal length of an OOS period has been based on experts' advice. Coolblue's restock management team specialists expect an OOS effect in the sales when a product is at least two weeks OOS. Furthermore, the second assumption ensures sufficient sales data to investigate.

Moreover, when we investigate OOS period j for product i, we will compare the sales data on days $\mathcal{T}_i^j$ and $\mathcal{T}_i^{j+1}$. When doing this, we will make the datasets of the same length by making the longest dataset as long as the shortest one. If $\mathcal{T}_i^j$ is the longest dataset, the first data points of $\mathcal{T}_i^j$ are removed, whereas if $\mathcal{T}_i^{j+1}$ is the longest dataset, the last data points of $\mathcal{T}_i^{j+1}$ are removed such that they become the same length. Let $T_i^{j,a}$ be the final length of the two datasets, this equals: $T_i^{j,a} = \min\{T_i^j, T_i^{j+1}\}$.

To clarify the assumptions, we will share some examples of products that meet and do not meet the assumptions, see Example 4.3.1.

Example 4.3.1: Products that meet and do not meet the assumptions of an OOS period

$$st_1^* = (PA, PA, PA, PA, PA, PA, PA, PU, PU, PU, PU, PU, PU, PU, PU, PU, PU, PU, PU, PU, PU, PA, PA, PA, PA, PA, PA, PA).$$

Product 1 meets all the assumptions: the OOS period has a length of 14 days and the data before and after the OOS period have a length of seven days. Therefore this product will be investigated.

$$st_2^* = (PA, PA, PA, PU, PU, PU, PU, PU, PA, PA, PA, PA, PA, PA, PU, PU, PA, PA, PA, PA, PA, PA, PA).$$

Product 2 has two OOS periods, both are not long enough and will be removed. A set of days with status-2 partial available is left, we will not investigate this product.

$$st_3^* = (PA, PA, PA, PU, PU, PU, PU, PU, PU, PU, PU, PU, PU, PU, PU, PU, PU, PU, PA, PA, PA, PA, PA, PA, PA, PA, PA, PA).$$

Product 3 has a long enough OOS period. However, there are only three days before the OOS period. Therefore, this product cannot be investigated.

$$st_4^* = (PA, PA, PA, PA, PU, PU, PA, PA, PA, PU, PU, PU, PU, PU, PU, PU, PU, PU, PU, PU, PU, PU, PU, PA, PA, PA, PA, PA, PA, PA, PA).$$

Product 4 has two OOS periods, the first one has a length of two days which is too short, thus it will be removed. Then, there is one long enough OOS period left with enough data before and after the OOS period. Therefore, we will investigate product 4. Moreover, the data before the OOS period has length seven, whereas the data after the OOS period has length eight. We need to equalise them by removing the last day of the data after the OOS period.

To conclude, products 2 and 3 are not being investigated, whereas products 1 and 4 will be investigated. However, the first OOS period and the last day in product 4 is being removed.

We have clarified our assumptions to investigate OOS periods, the next step is investigating them. We do this by describing perspectives about the sales data before and after OOS periods, these are sets of assumptions with hypotheses to be tested. There are two classes of perspectives: time-independent and dependent perspectives. In Section 4.4 we share the time-independent perspectives and in Section 4.5 the time-dependent perspectives.

4.4. Time-independent perspectives

Time-independent perspectives describe tests that investigate the sales data from a time-independent view: The data is assumed to be identical and independent distributed (iid). The perspectives are described per product, thus we omit the subscript for a product. For example, s_{it} becomes s_t. Moreover, we will describe the perspectives for products with one OOS period first, then we give the generalised version of these perspectives for products with multiple OOS periods.

4.4.1. Perspectives for products with one OOS period

We describe five perspectives for products with one OOS period: The first three Perspectives (4.4.1, 4.4.2, 4.4.3) are non-parametric while the last two Perspectives (4.4.4, 4.4.5) are parametric. All the products have one OOS period, therefore we will denote t_b^j and t_a^j by t_b and t_a, respectively, and $T^{j,a}$ becomes T^a.

To begin, we start with a simple perspective: We test if the distributions of the data before and after an OOS period are the same, see Perspective 4.4.1. The Kolmogorov-Smirnov (K-S) test [4] is used for this.

Perspective 4.4.1: Difference in distributions

Assume $\exists\ F_1,\ F_2$ such that $\forall\ t \leq t_b$ and $\forall\ t' \geq t_a$: $\quad s_t \overset{iid}{\sim} F_1, \quad s_{t'} \overset{iid}{\sim} F_2$:

- $H_0\ :\ F_1 = F_2$,
- $H_1\ :\ F_1 \neq F_2$.

We continue with a more concrete perspective by restricting the alternative hypothesis and testing for differences in probability, see Perspective 4.4.2. From this perspective, we can investigate which dataset is larger than the other in probability.

Perspective 4.4.2: Difference in probability

Assume $\exists\ \eta \in [0.5, 1]$, such that $\forall\ t \leq t_b$ and $\forall\ t' \geq t_a$:

- $H_0\ :\ s_t \overset{d}{=} s_{t'}$,
- $H_{1.a}\ :\ \mathbb{P}(s_t > s_{t'}) > \eta$,
- $H_{1.b}\ :\ \mathbb{P}(s_t < s_{t'}) > \eta$.

To test this perceptive, we link it to another concept, the one of stochastic dominance, see Definition 4.4.1 [8].

Definition 4.4.1: Stochastic dominance

Let X and Y be two random variables with continuous cumulative distribution functions F_X and F_Y, respectively. X *stochastically dominates* Y, denoted by $X \geq_{st} Y$, if

- $F_X(x) \leq F_Y(x) \qquad \forall\ x \in \mathbb{R}$,
- $F_X(x_0) < F_Y(x_0) \qquad \exists\ x_0 \in \mathbb{R}$.

Perspective 4.4.2 and Definition 4.4.1 are related because stochastic ordering implies inequality in probability as shown in Theorem 4.4.1.

Theorem 4.4.1: Stochastic dominance implies difference in probability

Let X and Y be two independent random variables with continuous cumulative distribution functions F_X and F_Y, respectively. If $X \geq_{st} Y$, then $P(X > Y) > \eta$, for some $\eta \in [0.5, 1]$.

Proof. We know by continuity that $\exists\ x_0, x_1, x_2$ with $x_1 < x_0 < x_2$ such that $F_X(x) < F_Y(x),\ \forall\ x \in (x_1, x_2)$. Then,

$$\begin{aligned} P(X > Y) = E_X(F_Y(X)) &= \int_{-\infty}^{\infty} F_Y(x) dF_X(x) \\ = \int_{-\infty}^{x_1} F_Y(x) dF_X(x) &+ \int_{x_1}^{x_2} F_Y(x) dF_X(x) + \int_{x_2}^{\infty} F_Y(x) dF_X(x) \\ > \int_{-\infty}^{x_1} F_X(x) dF_X(x) &+ \int_{x_1}^{x_2} F_X(x) dF_X(x) + \int_{x_2}^{\infty} F_X(x) dF_X(x) \\ &= \int_{-\infty}^{\infty} F_X(x) dF_X(x) = E_X(F_X(X)). \end{aligned}$$

Using the continuity and the probability integral transform ($Z = F_X(X) \rightarrow Z \sim U(0, 1)$):

$$P(X > Y) = E_X(F_Y(X)) > E_X(F_X(X)) = E_X(Z) = 0.5.$$

Thus, if $X \geq_{st} Y$, then $P(X > Y) > \eta$, for some $\eta \in [0.5, 1]$.

Theorem 4.4.1 also holds for discrete distributed random variables because these can be approximated by continuous distributions [4]. Therefore, Theorem 4.4.1 allows us to test Perspective 4.4.2 if we show stochastic ordering . The Wilcoxon-Mann-Whitney (WMW) test [26] can show this. Thus, Perspective 4.4.2 is tested by the WMW test.

The following perspective is less abstract than Perspective 4.4.2. In Perspective 4.4.3 the means of the datasets will be compared. The independent two-sample t-test [5] is used to do this. Note that one of the standard assumptions of the t-test is that the data is normally distributed. With our data, this assumption cannot be met for every product. However, the t-test also holds if not all the assumptions are met. See for example, Fay et al. (2010) [7], they show that the t-test is still asymptotically valid for more general assumptions.

Perspective 4.4.3: Difference in means

Assume $\exists\ \beta_1, \beta_2 \in \mathbb{R}$ such that $\forall\ t \leq t_b,\ s_t = \beta_1 + \epsilon_t$ with $\epsilon_t \overset{iid}{\sim} \mathbb{P}_1$ and $\forall\ t' \geq t_a,\ s_{t'} = \beta_2 + \epsilon_{t'}$ with $\epsilon_{t'} \overset{iid}{\sim} \mathbb{P}_2$, where $\mathbb{P}_1$ and $\mathbb{P}_2$ have a mean equal to zero and a finite fourth moment:

- $H_0 :\ \beta_1 = \beta_2$,
- $H_1 :\ \beta_1 \neq \beta_2$.

To show the differences between the three perspectives that just got introduced, we detail a short example. We have simulated an example product with one OOS period, see Figure 4.2. We created $s_t = (X_1, ..., X_{20}, U_1, ..., U_{15}, Y_1, ..., Y_{20})$ with $X_i \sim Pois(6)$, $U_i = 0$ and $Y_i \sim Pois(4)$. Perspectives 4.4.1, 4.4.2 and 4.4.3 test this data, with $s_t = (X_1, ..., X_{20})$ and $s_{t'} = (Y_1, ..., Y_{20})$. See Table 4.1 for the results. The K-S test is not able to find significant different distributions between the datasets, while the t-test and the WMW test can find significant different means and difference in probability, respectively.

Results of the example product

Persp.	Alt. hyp.	P-value	Test
4.4.1	1	0.18	K-S test
4.4.2	1.a	1.67e-02	WMW test
4.4.2	1.b	0.98	WMW test
4.4.3	1	1.95e-02	T-test

Table 4.1: Results of the example product in Figure 4.2.

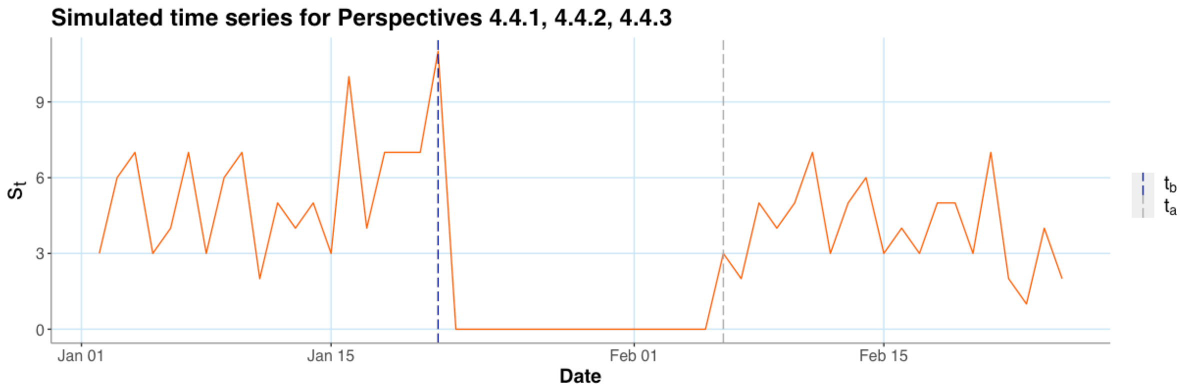

Figure 4.2: Simulated time series to show the difference in Perspectives 4.4.1, 4.4.2 and 4.4.3. The blue and grey dotted lines are t_{jb} and t_{ja}, the last day before and the first day after OOS period j, respectively.
We simulated $s_t = (X_1, ..., X_{20}, U_1, ..., U_{15}, Y_1, ..., Y_{20})$ with $X_i \sim Pois(6)$, $U_i = 0$ and $Y_i \sim Pois(4)$.

However, this is not coherent since these conclusions cannot hold simultaneously. The distributions should also be different if the means of the two datasets are different. One of the tests could give a wrong result since the probability of committing a type I or type II error is always strictly larger than zero. In this case, the K-S test used for Perspective 4.4.1 has made a type II error: It falsely does not reject the null hypothesis.

Another way of viewing the type II error is by looking at the power of the test. The power of a statistical test measures the frequency of not having a type II error. It is dependent on the type of test, the significance level, the sample size and the distribution of the data [22]. The types of tests are different in our example. In general, parametric tests (t-test) are more powerful than their non-parametric (K-S test) counterparts because non-parametric tests use less information in their calculation [15]. The assumptions of a parametric test help increase the test's power. However, if the assumptions of parametric tests are not met, the power can become smaller than the power of non-parametric tests. Wadgave et al. (2019) [25] gives examples of non-parametric tests being more powerful when the parametric assumption of normality is not satisfied.

Moreover, we have introduced the non-parametric perspectives, we only need to introduce the parametric perspectives. In the first non-parametric perspective, Perspective 4.4.4, we assume that the data before and after an OOS period follows a Poisson distribution, and we compare the parameters. The choice of a Poisson distribution is based on the fact that the sales data consists only of integers. As a result, we can assume the data follows a discrete distribution.

To test the perspective, we use test statistic $TS_{1,T^a} = \sqrt{\frac{T^a}{2\hat{\lambda}_a}}(\hat{\lambda}_2 - \hat{\lambda}_1)$ with $\hat{\lambda}_1$ and $\hat{\lambda}_2$ as the estimates of λ_1 and λ_2 respectively, $\hat{\lambda}_a = \frac{\hat{\lambda}_1+\hat{\lambda}_2}{2}$ and T^a the sample size of s_t and $s_{t'}$. Together with the two-sided Z-score table, we get a p-value for this test. See Section D.1 for the derivations of this test statistic.

Perspective 4.4.4: Different Poisson distributions

Assume $\exists\ \lambda_1, \lambda_2 > 0$ such that $\forall\ t \leq t_b,\quad s_t \overset{iid}{\sim} Pois(\lambda_1)$ and $\forall\ t' \geq t_a,\quad s_{t'} \overset{iid}{\sim} Pois(\lambda_2)$:

- $H_0:\ \lambda_1 = \lambda_2,$
- $H_1:\ \lambda_1 \neq \lambda_2.$

We continue to the last perspective of the time-independent perspectives, this perspective is applicable for a subset of the products. Perspective 4.4.5 is described for slow-moving products: These are products with a maximum of one sale per day. As a result, we assume that the sales data follows a Bernoulli distribution. Thus, Perspective 4.4.5 is for products for which $\max\{s_t\} = 1$ and it is tested with the Binomial test [1].

Perspective 4.4.5: Different Bernoulli distributions

Assume $\exists\ p_1, p_2 \in [0,1]$ such that $\forall\ t \leq t_b,\quad s_t \overset{iid}{\sim} Bernoulli(p_1)$ and $\forall\ t' \geq t_a,\quad s_{t'} \overset{iid}{\sim} Bernoulli(p_2)$:

- $H_0:\ p_1 = p_2,$
- $H_1:\ p_1 \neq p_2.$

4.4.2. Perspectives for products with multiple OOS periods

We continue to the perspectives for products with multiple OOS periods. These perspectives are a generalisation of the perspectives in Section 4.4.1. The perspectives investigate all the OOS periods of one product in different ways by describing multiple alternative hypotheses, we will introduce these alternative hypotheses by an example. In Figure 4.3 we have created multiple time series that show the different hypotheses for Perspective 4.4.6 (different distributions). For the null hypothesis, all four distributions of the in-stock sales data are the same. For hypothesis $1.a$, there is at least one data frame with a different distribution. For hypothesis $1.b$, all the consecutive data frames have different distributions; note that the distributions of data frames one and three, for example, still can be the same. Lastly, for hypothesis $1.c$, all the data frames have different distributions.

We will describe the time-independent perspectives for products with multiple OOS periods. As mentioned earlier, these are a generalisation of the perspectives in Section 4.4.1, therefore we will not go too detailed into the perspectives.

Perspective 4.4.6 tests if the distributions of the data before/after/in between the OOS periods are different. To test this, multiple K-S test are performed. For hypotheses $H_{1.a}$ and $H_{1.b}$ a Bonferroni

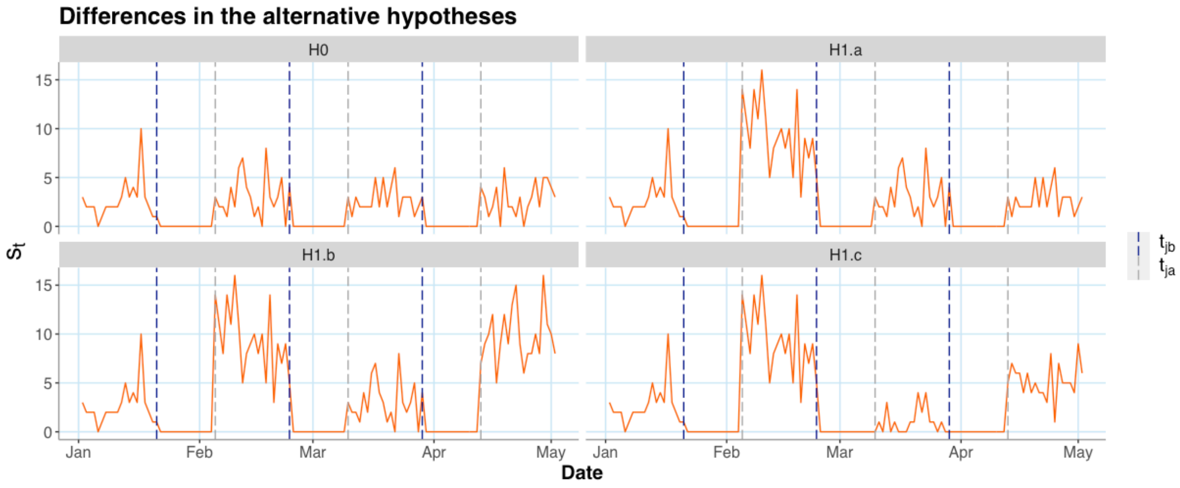

Figure 4.3: Visualisation of the different alternative hypotheses for Perspective 4.4.6. The blue and grey dotted lines are t_{jb} and t_{ja}, the last day before and the first day after OOS period j, respectively..
We simulated for $H_0 : s_t = (X_1, ..., X_{20}, U_1, ..., U_{14}, X_{21}, ..., X_{40}, U_{15}, ..., U_{28}, X_{41}, ..., X_{60}, U_{29}, ..., U_{42}, X_{61}, ..., X_{80})$,
for $H_{1.a} : s_t = (X_1, ..., X_{20}, U_1, ..., U_{14}, Y_1, ..., Y_{20}, U_{15}, ..., U_{28}, X_{21}, ..., X_{40}, U_{29}, ..., U_{42}, X_{41}, ..., X_{60})$,
for $H_{1.b} : s_t = (X_1, ..., X_{20}, U_1, ..., U_{14}, Y_1, ..., Y_{20}, U_{15}, ..., U_{28}, X_{21}, ..., X_{40}, U_{29}, ..., U_{42}, Y_{21}, ..., Y_{40})$,
and for $H_{1.c} : s_t = (X_1, ..., X_{20}, U_1, ..., U_{14}, Y_1, ..., Y_{20}, U_{15}, ..., U_{28}, Z_1, ..., Z_{20}, U_{29}, ..., U_{42}, W_1, ..., W_{20})$ with $X_i \sim Pois(3)$, $U_i = 0, Y_i \sim Pois(10)$, $Z_i \sim Pois(1)$ and $W_i \sim Pois(5)$.

correction [2] is applied, for hypothesis $H_{1.c}$ this is not applied.

Perspective 4.4.6: Difference in distributions

For $i = 1, ..., k+1$, assume $\exists\ F_i$ such that $(s_t)_{t \in \mathcal{T}^i} \overset{iid}{\sim} F^i$:

- $H_0 :\ \forall\ i, j \quad F^i = F^j$,
- $H_{1.a} :\ \exists\ i, j \quad F^i \neq F^j$,
- $H_{1.b} :\ \forall\ i \quad F^i \neq F^{i+1}$,
- $H_{1.c} :\ \forall\ i, j \quad F^i \neq F^j$.

We will explain why we apply a Bonferroni correction. If multiple hypothesis tests are being carried out simultaneously, the family-wise error rate (FWER) increases, this is the probability of making a type I error. The FWER is defined as:

$$FWER = 1 - (1 - \alpha)^m, \tag{4.1}$$

with α the α-level of an individual test and m the number of tests. For example, if we have a product with three OOS periods and we want to test hypothesis $H_{1.a}$ of Perspective 4.4.6, there are four different distributions to compare, and therefore there are six different tests. If $\alpha = 0.05$, our FWER equals 0.27. This means a 27% chance of having a type I error. To reduce this number, we can apply the Bonferroni correction:

$$\alpha_{new} = \frac{\alpha}{m}. \tag{4.2}$$

Our α_{new} becomes 0.008 if we apply the correction. With this new α-value our FWER equals 0.047, this is a much better value. Therefore, if we carry out these six tests, we test them with the K-S test with $\alpha = 0.008$.

Furthermore, if we want to test hypothesis $H_{1.b}$ of Perspective 4.4.6, we test three different tests simultaneously. Thus, our α-value becomes 0.017. For hypothesis $H_{1.c}$ of Perspective 4.4.6, we also test six different tests. However, for this hypothesis there is no need to control the FWER. Because we want to know if all the datasets are different.

Nevertheless, the Bonferroni correction also has its disadvantages. The Bonferroni correction can become too conservative, which means that the probability of type II errors increases if the number of tests gets large [14]. In our situation, the highest number of tests equals $\sum_{i=1}^{k_i} i$ for product i, we do not expect that k_i gets enormous, and therefore we do not expect that the Bonferroni correction gets too conservative.

We continue with describing the perspectives, Perspective 4.4.7 tests if the datasets are larger/smaller in probability. Also here, Theorem 4.4.1 is applied to use the WMW test. The WMW test is used with the Bonferroni correction for hypotheses $H_{1.a.a}, H_{1.a.b}, H_{1.b.a}$ and $H_{1.b.b}$, and without Bonferroni correct for hypotheses $H_{1.c.a}$ and $H_{1.c.b}$.

Perspective 4.4.7: Difference in probability

For $i, j = 1, ..., k+1$ assume $\exists\ \eta \in [0.5, 1]$ such that:

- $H_0 :\ \forall\ i, j \quad (s_t)_{t\in\mathcal{T}^i} \overset{d}{=} (s_{t'})_{t'\in\mathcal{T}^j}$,
- $H_{1.a.a} :\ \exists\ i, j \quad \mathbb{P}((s_t)_{t\in\mathcal{T}^i} > (s_{t'})_{t'\in\mathcal{T}^j}) > \eta$,
- $H_{1.a.b} :\ \exists\ i, j \quad \mathbb{P}((s_t)_{t\in\mathcal{T}^i} < (s_{t'})_{t'\in\mathcal{T}^j}) > \eta$,
- $H_{1.b.a} :\ \forall\ i \quad \mathbb{P}((s_t)_{t\in\mathcal{T}^i} > (s_{t'})_{t'\in\mathcal{T}^{i+1}}) > \eta$,
- $H_{1.b.b} :\ \forall\ i \quad \mathbb{P}((s_t)_{t\in\mathcal{T}^i} < (s_{t'})_{t'\in\mathcal{T}^{i+1}}) > \eta$,
- $H_{1.c.a} :\ \forall\ i, j \quad \mathbb{P}((s_t)_{t\in\mathcal{T}^i} > (s_{t'})_{t'\in\mathcal{T}^j}) > \eta$,
- $H_{1.c.b} :\ \forall\ i, j \quad \mathbb{P}((s_t)_{t\in\mathcal{T}^i} < (s_{t'})_{t'\in\mathcal{T}^j}) > \eta$.

The third generalised perspective compares the means of the datasets, see Perspective 4.4.8. The t-test with Bonferroni correction is used for hypotheses $H_{1.a}$ and $H_{1.b}$ and without correction for hypothesis $H_{1.c}$.

Perspective 4.4.8: Difference in means

For $i = 1, ..., k+1$ assume $\exists\ \beta^i \in \mathbb{R}$, such that $\forall\ t \in \mathcal{T}^i :\ s_t = \beta^i + \epsilon_t^i$ with $\epsilon_t^i \sim \mathbb{P}^i$ where $\mathbb{P}^i$ has a mean equal to zero and a finite fourth moment:

- $H_0 :\ \forall\ i, j \quad \beta^i = \beta^j$,
- $H_{1.a} :\ \exists\ i, j \quad \beta^i \neq \beta^j$,
- $H_{1.b} :\ \forall\ i \quad \beta^i \neq \beta^{i+1}$,
- $H_{1.c} :\ \forall\ i, j \quad \beta^i \neq \beta^j$.

We have covered the non-parametric perspectives, we continue to the parametric perspectives. Perspective 4.4.9 tests if the data before/after/in between the OOS have the same Poisson parameters. This is being tested with the same test statistic from Perspective 4.4.4, but rewritten for multiple OOS periods: $TS_{1,T^{i,j,a},i,j} = \sqrt{\frac{T^{i,j,a}}{2\hat{\lambda}_{i,j,a}}}(\hat{\lambda}^j - \hat{\lambda}^i)$ with $\hat{\lambda}^i$ and $\hat{\lambda}^j$ the estimates of λ^i and λ^j respectively, $\hat{\lambda}_{i,j,a} = \frac{\hat{\lambda}^i + \hat{\lambda}^j}{2}$ and $T^{i,j,a}$ the length of $\mathcal{T}^i$ and $\mathcal{T}^j$. For hypotheses $H_{1.a}$ and $H_{1.b}$ the Bonferroni correction is applied, for hypothesis $H_{1.c}$ this is not done.

Perspective 4.4.9: Different Poisson distributions

For $i = 1, ..., k+1$, assume $\exists\ \lambda^i > 0$ such that $(s_t)_{t\in\mathcal{T}^i} \overset{iid}{\sim} Pois(\lambda^i)$:

- $H_0 :\ \forall\ i, j \quad \lambda^i = \lambda^j$,
- $H_{1.a} :\ \exists\ i, j \quad \lambda^i \neq \lambda^j$,
- $H_{1.b} :\ \forall\ i \quad \lambda^i \neq \lambda^{i+1}$,
- $H_{1.c} :\ \forall\ i, j \quad \lambda^i \neq \lambda^j$.

Finally, Perspective 4.4.10 tests if the data before/after/in between the OOS have the same Bernoulli parameter. To test this, multiple Binomial tests are used. For hypotheses $H_{1.a}$ and $H_{1.b}$ a Bonferroni correction is applied, for hypothesis $H_{1.c}$ this is not applied. Also, this perspective can only be used for products with $\max\{s_t\} = 1$.

Perspective 4.4.10: Different Bernoulli distributions

For $i = 1, ..., k+1,\ \exists\ p^i \in [0,1]:\ (s_t)_{t\in\mathcal{T}^i} \overset{iid}{\sim} Bernoulli(p^i)$:

- $H_0:\ \forall\ i,j \quad p^i = p^j$,
- $H_{1.a}:\ \exists\ i,j \quad p^i \neq p^j$,
- $H_{1.b}:\ \forall\ i \quad p^i \neq p^{i+1}$,
- $H_{1.c}:\ \forall\ i,j \quad p^i \neq p^j$.

All the time-independent perspectives are described for products with one OOS period and multiple products. Table E.2 in the Section E.2 summarises all the time-independent perspectives with a short description and their tests. The perspectives will test all the OOS periods that meet the assumptions introduced in Section 4.3 and give a percentage of how many OOS periods reject the hypothesis tests per perspective. These results will be shared in Section 4.7. We continue to the other class of perspectives: time-dependent perspectives.

4.5. Time-dependent perspectives

Another way to investigate the sales data is by fitting a time series on them and investigating those. The time-dependent perspectives do this; we use autoregressive moving average (ARMA) processes and integer-generalised autoregressive conditional heteroskedasticity (INGARCH) processes. The ARMA processes are used since these are one of the most well-known time series. However, ARMA processes are described for time series with rational numbers, whereas our data consists only of integers. Therefore, we also use INGARCH processes to fit our data since these are specially constructed for integer data.

First, we show that our sales data is time-dependent in Section 4.5.1. Then, we determine the optimal orders for our ARMA and INGARCH processes in Sections 4.5.2 and 4.5.3. With these orders, we describe the time-dependent perspectives in Sections 4.5.4 and 4.5.5. Also, these perspectives are described per product, therefore we omit the subscript for a product.

4.5.1. Dependence of the sales data

Before we describe time-dependent perspectives, we want to know if our data is time-dependent, otherwise there is no reason to describe time-dependent perspectives. We do this by fitting an $AR(p)$ process on all the in-stock data of a product. See Definition 4.5.1 [3] for the definition of an $AR(p)$ process. Note that there are more methods to show time-dependence in data than by fitting an $AR(p)$ process.

Definition 4.5.1: $AR(p,q)$ process

X_t is an $AR(p,q)$ process, if $\exists\ p > 0$ and $\phi_p \neq 0$:

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + Z_t, \qquad Z_t \sim WN(0, \sigma^2). \tag{4.3}$$

The model can be rewritten into:

$$\phi(B) X_t = Z_t, \tag{4.4}$$

with B the backward shift operator ($B^k X_t = X_{t-k}$), $\phi(\cdot)$ the p-th degree polynomials:

$$\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p. \tag{4.5}$$

To fit an $AR(p)$ process, we use the **ar()** function from the standard R library **stats**. It fits an $AR(p)$ process and estimates the order p using the AIC. In the **ar()** function, we set the argument "order.max" equal to $\sum_{j=1}^{k+1} T^j - 1$. The input data are the sales when the status-2 equals partial available, thus we fit the function on $(s_t)_{t \in \mathcal{T}^{\text{partial available}}}$. Then, for every product, we investigate the value of the estimation of p. There is no dependence in the time series if $p = 0$, whereas there is dependence if $p > 0$. Note that for this investigation, we do not use the assumptions given in Section 4.3, which means that we use all the in-stock sales data despite of the lengths of OOS periods.

The data used for this investigation is the product information from 12/21/2019 until 12/20/2021 of products sold in the Netherlands. This is the same data frame used in the previous chapters but extended by one year, such that we have more OOS periods to investigate. For the other investigations and results in this chapter, this same 2-year product information data set is used.

Author's note: This figure is confidential..

Figure 4.4: Results of the time-dependence investigation.

Author's note: This paragraph is confidential..

4.5.2. Order of the $ARMA(p,q)$ processes

To use $ARMA(p,q)$ processes in perspectives, we should determine the optimal order, this are the values of p and q. First, we share the definition of an $ARMA(p,q)$ process, see Definition 4.5.2 [3].

Definition 4.5.2: $ARMA(p,q)$ process

X_t is a stationary $ARMA(p,q)$ *process*, if $\exists\ p,q > 0$, $\sigma^2 \neq 0$, $\phi_p \neq 0$ and $\theta_q \neq 0$:

$$X_t - \sum_{k=1}^{p} \phi_k X_{t-k} = Z_t + \sum_{l=1}^{q} \theta_l Z_{t-l}, \qquad Z_t \sim WN(0, \sigma^2), \tag{4.6}$$

where $(1 - \phi_1 z - ... - \phi_p z^p)$ and $(1 + \theta_1 z + ... + \theta_q z^q)$ have no common factors.
The model can we rewritten into:

$$\phi(B) X_t = \theta(B) Z_t, \tag{4.7}$$

with B the backward shift operator ($B^k X_t = X_{t-k}$), $\phi(\cdot)$ and $\theta(\cdot)$ the p-th and q-th degree polynomials:

$$\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p, \tag{4.8}$$

and,

$$\theta(z) = 1 + \theta_1 z + ... + \theta_q z^q. \tag{4.9}$$

For stationary, the following should hold:

$$\phi(z) \neq 0 \qquad \text{for } z = \pm 1. \tag{4.10}$$

To determine the value of p and q, we will fit $ARMA(p,q)$ processes on the product's sales data when it has status partial available ($(s_t)_{t \in \mathcal{T}^{\text{partial available}}}$), for p and q ranging from zero till 15. We select 10,000 products that meet the following criteria:

1. Have order $p > 0$ from the time-dependence investigation in Section 4.5.1, else we are optimising p and q on time-independent data.
2. Have sufficient amount of data to fit all the processes on ($|\mathcal{T}^{\text{partial available}}| \geq 16$), else is fitting an $ARMA(p,q)$ process with high orders impossible.
3. Have non-constant data ($\min(s_t) \neq \max(s_t)$), else we cannot fit an $ARMA(p,q)$ process.

With these products, we calculate the AIC and BIC for every combination of p and q of the fitted $ARMA(p,q)$ process for every product i:

$$AIC_{i,p,q} = -2\ln(L_{i,p,q}) + 2(p+q+2), \qquad p,q = 0,...,15, \tag{4.11}$$

$$BIC_{i,p,q} = -2\ln(L_{i,p,q}) + \ln(T_i^{\text{partial available}})(p+q+2), \qquad p,q = 0,...,15, \tag{4.12}$$

with n the number of products, $L_{i,p,q}$ the likelihood of the fitted $ARMA(p,q)$ process on product i and $T_i^{\text{partial available}}$ the length of the dataset of product i. The number of parameters equals $p+q+2$ since we have $p+q$ coefficients, the variance and the intercept. A lower AIC or BIC indicates a better model. We use the **arima()** function from the **stats** library to fit $ARMA(p,q)$ processes. In the **arima()** function, we choose the method "CSS", which means that the conditional sum-of-squares are minimised. Unfortunately, sometimes the **arima()** function fails to fit a model for a given p and q due to different reasons. Table E.1 in Section E.1 gives an overview of the amount of time the function failed.

For all combinations of p and q, $\upsilon_{AIC,p,q}$ and $\upsilon_{AIC,p,q}$ are calculated, see Equations 4.13 and 4.14. These percentages tell us for how many products the $ARMA(p,q)$ is chosen as optimal fit based on the AIC and BIC over all the successful $ARMA(p,q)$ fits.

$$\upsilon_{AIC,p,q} = \frac{\sum_{i=1}^{10,000} \mathbb{1}_{\{AIC_{i,p,q} = \min_{\forall\, p,\, q}(AIC_{i,p,q})\}}}{\text{\# products for which the } ARMA(p,q) \text{ fit succeeded}} * 100 \tag{4.13}$$

$$\upsilon_{BIC,p,q} = \frac{\sum_{i=1}^{10,000} \mathbb{1}_{\{BIC_{i,p,q} = \min_{\forall\, p,\, q}(BIC_{i,p,q})\}}}{\text{\# products for which the } ARMA(p,q) \text{ fit succeeded}} * 100 \tag{4.14}$$

Figure 4.5 is a heat map of all values of $\upsilon_{AIC,p,q}$ and $\upsilon_{AIC,p,q}$ for $p,q = 0,...,15$. The AIC has the highest percentage at $p = q = 15$, whereas the BIC has the highest percentage at $p = q = 1$. The BIC metric's result is more convincing than the AIC metric's result since the percentage is much higher. However, these results are still contradicting, we will discuss the two metrics to decide which values of p and q are optimal.

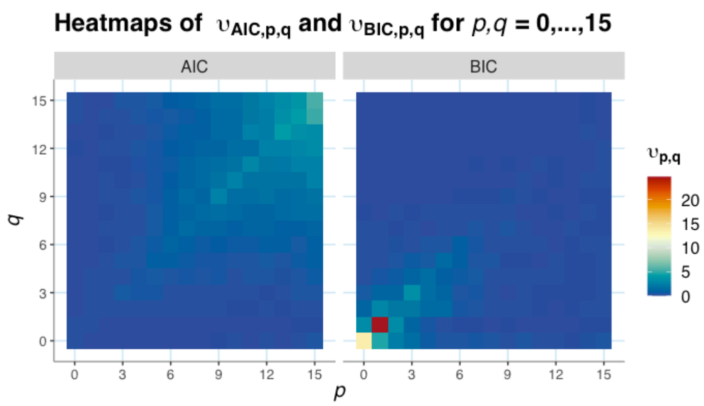

Figure 4.5: Heatmaps of $\upsilon_{AIC,p,q}$ and $\upsilon_{BIC,p,q}$ for $p,q = 0,...,15$, the percentages per $ARMA(p,q)$ fit that is selected as the best fit per product according to the AIC and BIC.

The AIC and BIC formulas are nearly the same, the difference is the penalty of the number of parameters. For AIC, the penalty equals $2k$, whereas for BIC, the penalty equals $ln(n)k$, with k the number of parameters and n the sample size. In Figure 4.5 we see that the penalty $2k$ in the AIC does not severely penalise complex models since the complex models have the best results. Whereas the $ln(n)k$ penalty in the BIC does penalise the complex models more heavily, the BIC has the worst results for higher orders.

Furthermore, if a true model in the families of the $ARMA(p,q)$-models exists for the sales data of a product, the BIC is consistent, and it will select the true model asymptotically. On the contrary, if there is no true model for the data, the AIC is efficient, and it will select the best-approximating model asymptotically [24]. We will have products' sales data for which a true model exists, but we will also have products' sales data for which no true model exists. Thus our data is a mixture of existing and non-existing true models; we cannot make a conclusion based on this characteristic of the metrics.

Another result we got earlier is the number of failures of fitting an $ARMA(p,q)$ model with the **arima()** function. In Table E.1 we see that for $p = q = 1$, 2.80% of the products the function failed whereas for $p = q = 15$, 71.30% of the products the function failed. This means that the likelihood of a failure is larger when we set $p = q = 15$ than for $p = q = 1$.

Based on the last argument, the results of the BIC and the simplicity of computations that we have to do later in the perspectives, we choose $p = q = 1$.

4.5.3. Order of the $INGARCH(p,q)$ processes

We also want to determine the optimal order for $INGARCH(p,q)$ processes in the same way we did for $ARMA(p,q)$ processes. First, we introduce $INGARCH(p,q)$ processes and our motivation to use them, then we elaborate on our choice of p and q.

When we use $ARMA(1,1)$ processes to describe our data, time series with rational numbers will be fitted. This does not match our data since we have count time series, these are time series that consist of non-negative integers. Therefore, we can use this characteristic to fit another process. We will use a Poisson distribution to describe the data as an integer-valued generalised autoregressive conditional heteroskedasticity process ($INGARCH$). See Definition 4.5.3 for the definition of an $INGARCH(p,q)$ process.

Definition 4.5.3: $INGARCH(p,q)$ process

Let $\mathcal{F}_t = \sigma(s_i,\ i \leq t)$ denote the history of the process s_t, $\alpha > 0$ a constant and $\sum_{j=1}^{p} \phi_j + \sum_{j=1}^{q} \theta_j < 1$ coefficients with $\phi_1, ..., \phi_p, \theta_1, ..., \theta_q \geq 0$. Then for $p, q > 0$, X_t is a stationary $INGARCH(p,q)$ *process* if

$$s_t|\mathcal{F}_{t-1} \sim Pois(\lambda_t), \qquad \lambda_t = \alpha + \sum_{k=1}^{p} \phi_k s_{t-k} + \sum_{l=1}^{q} \theta_l \lambda_{t-l}. \tag{4.15}$$

For stationary, the following should hold:

$$\sum_{k=1}^{p} \phi_k + \sum_{l=1}^{q} \theta_l < 1 \tag{4.16}$$

We use the the **tsglm()** function of the **tscount** library [12] to fit an $INGARCH(p,q)$ process. We wish to optimise p and q in the same ways as we did in Section 4.5.2 for $ARMA(p,q)$ processes. However, this is impossible because the function **tsglm()** has a long computation time: Fitting the function for $p,q = 1, ..., 6$ for one product takes approximately 3.5 minutes. If we want to do this for $p,q = 0, ..., 15$ for 10,000 products, it will take approximately 173 days to run. This takes too much time. As a result, we choose $p = q = 1$ because this saves complex computations when using the $INGARCH(p,q)$ process in the perspectives. Also, we expect that for higher orders, the function will frequently fail, the same as with the **arima()** function. Therefore we choose a lower order.

We have introduced $ARMA(p, q)$ and $INGARCH(p, q)$ processes and concluded that the order $p = 1$ and $q = 1$ are optimal for both time series. Also, we saw that not every product's sales data could be fitted by one of these processes. Therefore, using these processes in perspectives does not mean that every OOS period can be investigated and that every fit has a high quality.

4.5.4. Perspectives for products with one OOS period

We start with introducing two perspectives in which $ARMA(1, 1)$ processes are used, after that we introduce three perspectives in which $INGARCH(1, 1)$ processes are used. Note that for this section we describe the perspectives for products with one OOS period, therefore we will rewrite t_{jb} and t_{ja} as t_b and t_a, respectively.

In Perspective 4.5.1 we want to test if the sales at t_a is an outlier since this is the first day after the OOS period. If an OOS period influences the sales, we expect at t_a an abnormal decrease or increase in the sales. We do this with two different methods: In the first method we fit an $ARMA(1, 1)$ process to $(s_t)_{t\in(\mathcal{T}^1,\mathcal{T}^2)}$ and investigate the residual on time t_a, in the second method we fit an $ARMA(1, 1)$ process to $(s_t)_{t\in\mathcal{T}^1}$ and predict the sales on t_a, then we investigate this prediction. We test Perspectives 4.5.1.1 and 4.5.1.2 using the test statistics $TS_2 = \frac{\epsilon_{t_a}}{\hat{\sigma}}$ and $TS_3 = \frac{\epsilon^*_{t_a}}{\hat{\sigma}^*}$, respectively, with $\hat{\sigma} = sd(\epsilon_t)$ and $\hat{\sigma}^* = sd(\epsilon^*_t)$. Together with the two-sided Z-score table we get a p-value for these tests.

Perspective 4.5.1: The value s_{t_a} is an outlier

- Perspective 4.5.1.1: Difference between $\bar{s}_{t_a}$ and s_{t_a}

Let $(s_t)_{t\in(\mathcal{T}^1,\mathcal{T}^2)} \sim ARMA(1, 1)$ with fitted values $\bar{s}_t$, residuals $\epsilon_t = \bar{s}_t - s_t$ and $\sigma = sd(\epsilon_t)$. Let c be a positive constant.

- $H_0 : \frac{\epsilon_{t_a}}{\sigma} \sim N(0, 1)$,
- $H_1 : \left|\frac{\epsilon_{t_a}}{\sigma}\right| > c.$

- Perspective 4.5.1.2: Difference in $\hat{s}_{t_a}$ and s_{t_a}

Let $(s_t)_{t\in\mathcal{T}^1} \sim ARMA(1, 1)$ with fitted values $\bar{s}_t$, predicted values $\hat{s}_t$, residuals $\epsilon_t = \bar{s}_t - s_t$ and $\epsilon^*_t = \hat{s}_t - s_t$, and $\sigma^* = sd(\epsilon^*_t)$. Let c be a positive constant.

- $H_0 : \frac{\epsilon^*_{t_a}}{\sigma^*} \sim N(0, 1)$,
- $H_1 : \left|\frac{\epsilon^*_{t_a}}{\sigma^*}\right| > c.$

We only investigate the residual at time t_a because the influence of an outlier in the residuals of the fitted time series is only perceptible at that time point t_a. We will demonstrate this with an example. Figure 4.6 shows an example time series with the fitted values of an $ARMA(1, 1)$ process and its residuals. There is an outlier on the 20th of February, this is also visible in the residuals, but only on that date. The $ARMA(1, 1)$ process needs one time point to adjust to the outlier, therefore we only investigate the sales at t_a in Perspective 4.5.1.

We continue to the second perspective with an $ARMA(1, 1)$ process. In this perspective, we do not investigate one time point, but we compare the data before and after the OOS period by fitting two separate $ARMA(1, 1)$ processes. One is fitted to $(s_t)_{t\in\mathcal{T}^1}$ and one to $(s_t)_{t\in\mathcal{T}^2}$, see Perspective 4.5.2. We compare the coefficients of the two processes with a special distance function, see Definition 4.5.4 [17].

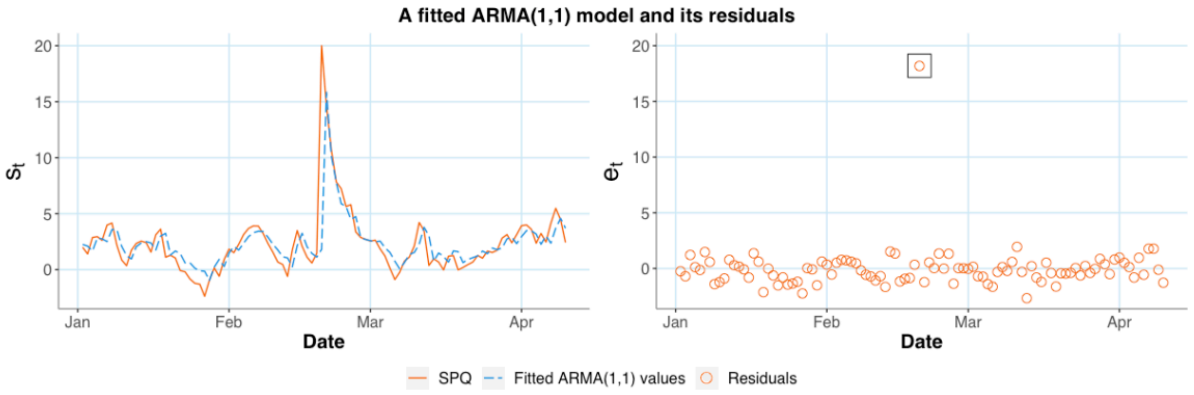

Figure 4.6: An example of time series with a fitted $ARMA(1,1)$ process and its residuals. On the 20th of February there is an outlier, the residual on this date is marked with a square.

We simulated $s_t = 0.4 + 0.7s_{t-1} + Z_t + 0.3Z_{t-1}$, for $t = 1, ..., 100$ with $s_1 = 2$, $s_{50} = 20$ and $Z_t \overset{iid}{\sim} N(0, 1)$.

Perspective 4.5.2: Difference in coefficients between two $ARMA(1,1)$ processes

Let $(s_t)_{t\in\mathcal{T}^1} \sim ARMA(1,1)$ and let $(s_{t'})_{t'\in\mathcal{T}^2} \sim ARMA(1,1)$.

- $H_0 : d((s_t)_{t\in\mathcal{T}^1}, (s_{t'})_{t'\in\mathcal{T}^2}) = 0,$
- $H_1 : d((s_t)_{t\in\mathcal{T}^1}, (s_{t'})_{t'\in\mathcal{T}^2}) \neq 0,$

with $d(\cdot,\cdot)$ the distance metric from Definition 4.5.4.

Definition 4.5.4: Distance between two $ARMA(p,q)$ processes

The *distance between* $ARMA(p,q)$ processes X_t and Y_t is defined as,

$$d(X_t, Y_t) = \sqrt{\sum_{j=1}^{\infty} (\pi_{j,x} - \pi_{j,y})^2} = \sqrt{(\boldsymbol{\pi}_x - \boldsymbol{\pi}_y)^T (\boldsymbol{\pi}_x - \boldsymbol{\pi}_y)}.$$

For this definition, we introduced a new variable, variable π. The π-values come from the $ARMA(p,q)$ coefficients:

$$\pi(z) = \frac{\phi(z)}{\theta(z)}, \tag{4.17}$$

with $\pi(z) = 1 + \sum_{j=1}^{\infty} \pi_j z^j$, the coefficients can be written as a vector $\boldsymbol{\pi} = (\pi_1, ..., \pi_j, ...)^T$. They are also known as the coefficients of the $AR(\infty)$ operator. For an $ARMA(1,1)$ process the distance becomes:

$$d^2(X_t, Y_t) = \frac{(\phi_x + \theta_x)^2}{1 - \theta_x^2} + \frac{(\phi_y + \theta_y)^2}{1 - \theta_y^2} - 2\frac{(\phi_x + \theta_x)(\phi_y + \theta_y)}{1 - \theta_x\theta_y}. \tag{4.18}$$

In Section D.2 are the calculations for Equation 4.18. Moreover, we test Perspective 4.5.2 using the test statistic created by Maharaj (1996) [13], see Theorem 4.5.1. For this test statistic, the infinite order of the AR coefficients is truncated into $k = p + q = 2$, thus we have $\boldsymbol{\pi} = (\pi_1, \pi_2)^T$. Equation 4.19 follows a chi-square distribution with k degrees of freedom. Thus, using test statistic $TS_{4,\mathcal{T}^a}$ together with the chi-square distribution table we can test Perspective 4.5.2. We have written out some elements of the test statistic $TS_{4,\mathcal{T}^a}$ for an $ARMA(1,1)$ process in Section D.3 in the Appendix.

Theorem 4.5.1: Test statistic for the $ARMA(p,q)$ distance metric

$$TS_{4,T^a} = T^a(\hat{\boldsymbol{\pi}}_x - \hat{\boldsymbol{\pi}}_y)^T \hat{V}^{-1}(\hat{\boldsymbol{\pi}}_x - \hat{\boldsymbol{\pi}}_y) \overset{A}{\sim} \chi^2(k), \tag{4.19}$$

with,

- $T^a = T^1 = T^2$,
- $\hat{\boldsymbol{\pi}}_x$ and $\hat{\boldsymbol{\pi}}_y$ are the estimates of $\boldsymbol{\pi_x}$ and $\boldsymbol{\pi_y}$, respectively,
- $\hat{V} = (\hat{\sigma}_x^2 \hat{R}_x^{-1}(k) + \hat{\sigma}_y^2 R_y^{-1}(k))$ with $k = p + q$,
- $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$ are the estimated variances of the white noise ,
- $\hat{R}_x(k)$ and $\hat{R}_y(k)$ are the upper $k \times k$ sub-matrices of the estimated infinite auto-correlation matrices.

We proceed with the perspectives that use $INGARCH(1,1)$ processes. In Perspective 4.5.3 we do the same test as in Perspective 4.5.1, we test if s_{t_a} is an outlier with two different methods. Also here we use the test statistics $TS_2 = \frac{\epsilon_{t_a}}{\hat{\sigma}}$ and $TS_3 = \frac{\epsilon^*_{t_a}}{\hat{\sigma}^*}$ and the two-tailed Z-score table.

Perspective 4.5.3: The value s_{t_a} is an outlier

- Perspective 4.5.3.1: Difference between $\bar{s}_{t_a}$ and s_{t_a}

Let $(s_t)_{t \in (\mathcal{T}^1, \mathcal{T}^2)} \sim INGARCH(1,1)$ with fitted values $\bar{s}_t$, residuals $\epsilon_t = \bar{s}_t - s_t$ and $\sigma = sd(\epsilon_t)$. Let c be a positive constant.

- $H_0 :\ \dfrac{\epsilon_{t_a}}{\sigma} \sim N(0,1)$,
- $H_1 :\ \left|\dfrac{\epsilon_{t_a}}{\sigma}\right| > c$.

- Perspective 4.5.3.2: Difference in $\hat{s}_{t_a}$ and s_{t_a}

Let $(s_t)_{t \in \mathcal{T}^1} \sim INGARCH(1,1)$ with fitted values $\bar{s}_t$, predicted values $\hat{s}_t$, residuals $\epsilon_t = \bar{s}_t - s_t$ and $\epsilon^*_t = \hat{s}_t - s_t$, and $\sigma^* = sd(\epsilon^*_t)$. Let c be a positive constant.

- $H_0 :\ \dfrac{\epsilon^*_{t_a}}{\sigma^*} \sim N(0,1)$,
- $H_1 :\ \left|\dfrac{\epsilon^*_{t_a}}{\sigma^*}\right| > c$.

Furthermore, we also compare the data before and after the OOS period by fitting two separate $INGARCH(1,1)$ processes, see Perspective 4.5.4.

To test this perspective, we have created another test statistic. We use test statistic $TS_{5,T^a} = \left\|\sqrt{\frac{TG_{T^a}(\hat{\boldsymbol{\omega}}_a)}{2}}(\hat{\boldsymbol{\omega}}_1 - \hat{\boldsymbol{\omega}}_2)\right\|_2^2$ with the chi-square distribution table. In Section D.4 in the Appendix are the derivations and definitions of TS_{5,T^a}.

Perspective 4.5.4: Difference in coefficients between two $INGARCH(1,1)$ processes

Let $(s_t)_{t \in \mathcal{T}^1} \sim INGARCH(1,1)$ and let $(s_t)_{t \in \mathcal{T}^2} \sim INGARCH(1,1)$, and let $\boldsymbol{\omega}_1$ and $\boldsymbol{\omega}_2$ be the parameter vectors of the two $INGARCH(1,1)$ processes.

- $H_0 :\ \boldsymbol{\omega}_1 = \boldsymbol{\omega}_2$,
- $H_1 :\ \boldsymbol{\omega}_1 \neq \boldsymbol{\omega}_2$.

For the last perspective we use a function from the **tscount** library, it has a function that allows to include intervention effects in the $INGARCH(p,q)$ process, as a result an $INGARCH(1,1)$ process with intervention effect at t_a becomes:

$$s_t|\mathcal{F}_{t-1} \sim Pois(\lambda_t), \qquad \lambda_t = \alpha + \phi_1 s_{t-1} + \theta_1 \lambda_{t-1} + \eta_{t_a} \mathbb{1}_{\{t=t_a\}}, \tag{4.20}$$

with $\eta_{t_a} \geq 0$ the effect on t_a. In Perspective 4.5.5 we test if the intervention effect at t_a has an significant effect. For this we use the **interv_test()** function of the **tscount** library with argument delta equal to zero. The function tests if the fitted model by function **tsglm()** has any significant intervention effects at time t_a.

Perspective 4.5.5: Intervention effect of $INGARCH(1,1)$ process

Let $(s_t)_{t\in(\mathcal{T}^1,\mathcal{T}^2)} \sim INGARCH(1,1)$ with an intervention effect at time t_a.

- $H_0 : \eta_{t_a} = 0$,
- $H_1 : \eta_{t_a} \neq 0$.

4.5.5. Perspectives for products with multiple OOS periods

The time-dependent perspectives for products with multiple OOS periods are a generalisation of the perspectives introduced in the previous section. Therefore, we will not describe them in this section, but share them in Chapter F. Moreover, also for the time-dependent perspectives we have a summary, see Table E.3 in Section E.2.

We have shared all the time-independent and time-dependent perspectives that we will use. These perspectives test the products' OOS period(s) individually and show how many percentage of the products for which the perspective is tested accept or reject the hypothesis test. However, we can do more with these perspectives. We will introduce this in the next section.

4.6. Perspectives for multiple products

Another way to investigate the OOS periods is by analysing certain parameters that give information about all the OOS periods. For example, take the residual ϵ_{t_a} from Perspective 4.5.1.1 for every OOS period. We can investigate the distribution of this parameter: Is it skewed, and does it has a positive or negative mean? Investigating these parameters' distribution gives us results from another viewpoint. We will investigate parameters that are already introduced in perspectives before. First, we will share which parameters we are going to investigate. After that, the perspective used to investigate them is shared. From now on, we will call these parameters the measures of interest.

In Table 4.2 is given which measures of interest we want to investigate and from which perspective they come. Every measure of interest describes something from an OOS period j for product i. If OOS periods do not influence the sales, we expected that these measures equal zero.

Overview of the measures of interest

Measure of interest	Perspective
$\beta_i^{j+1} - \beta_i^j$	4.4.3 & 4.4.8
$\epsilon_{i,t_{ja}}$	4.5.1.1 & F.0.1.2
$\epsilon^*_{i,t_{ja}}$	4.5.1.2 & F.0.1.2
$d((s_t)_{t\in\mathcal{T}_i^j}, (s_t)_{t\in\mathcal{T}_i^{j+1}})^2$	4.5.2 & F.0.2
$\epsilon_{i,t_{ja}}$	4.5.3.1 & F.0.3.1
$\epsilon^*_{i,t_{ja}}$	4.5.3.2 & F.0.3.2
$\eta_{i,t_{ja}}$	4.5.5 & F.0.5

Table 4.2: An overview of the measures of interest that will be investigated, with $i = 1,...,n$ the product and $j = 1,...,k_i + 1$ the OOS period. Together with the perspective from which the measure comes.

We continue by describing the perspective such that we can investigate the measures of interest.

We describe one generalised perspective for a parameter θ_i^j, this perspective can be applied to every measure of interest. Let θ_i^j describe $(s_{it})_{t\in\{\mathcal{T}_i^j,\mathcal{T}_i^{j+1}\}}$ for product $i = 1, ..., n$ and $j = 1, ..., k_i + 1$, we will investigate θ_i^j for all i and j, see Perspective 4.6.1.

Perspective 4.6.1: Distribution of θ_i^j

For $i = 1, ..., n$ and $j = 1, ..., k_i$ we have random variables $\theta_i^j \in \mathbb{R}$ that describe every the time frame $\mathcal{T}_i^j$. Let $\sigma = sd(\theta_i^j)\ \forall\ i, j$. Then, we assume $\exists\ \mathbb{P}_{\theta_i^j}$ such that $s_{it}|\theta_i^j \overset{\perp}{\sim} \mathbb{P}_{\theta_i^j}$. Furthermore, we assume $\frac{\theta_i^j}{\sigma} \overset{iid}{\sim} F$ for some distribution F with standard deviation 1.

- Perspective 4.6.1.1: Mean of θ_i^j

- $H_0 : \mu(F) = 0,$
- $H_1 : \mu(F) \neq 0.$

- Perspective 4.6.1.2: Median of θ_i^j

- $H_0 :$ med$(F) = 0,$
- $H_1 :$ med$(F) \neq 0.$

Perspective 4.6.1.1 will be tested with the t-test and Perspective 4.6.1.2 with the WMW test. Note that we normalise every measure of interest before assuming that it follows some distribution F in the perspective. We do this such that the measures have a standard scale. We have chosen a specific method to normalise the measures, in Section 4.7.1 we will detail more about this choice.

We have introduced multiple new variables in Perspective 4.6.1. For clarity, we have made an overview of the type of variables, see Figure 4.7.

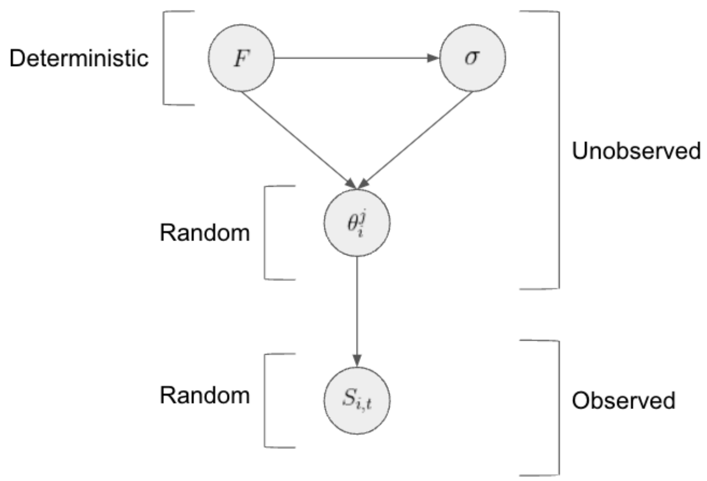

Figure 4.7: Overview of the variables of Perspective 4.6.1. $a \rightarrow b$ means that a is dependent on b.

4.7. Results

We have introduced all the perspectives we use to investigate if an OOS period influences sales. In this section, we will share the results of all the perspectives. First, we will share our motivation for choosing the normalisation method in Perspective 4.6.1. After that, we will share the results of the perspectives in the same order as we introduced them.

For the results, we use the two-year product information data set from 2019/12/21 until 2021/12/20 for products sold in the Netherlands. This data set is one year longer than the data sets used in the previous chapters. This is because we can investigate more OOS periods with a larger data set. The data set consists out of 70,646 different products.

4.7.1. Normalisation of the measures of interest

In Perspective 4.6.1 we normalise the measure we are describing such that all the measures have a common scale. There are multiple ways to normalise a variable. We have compared two choices of normalisation, we will introduce them by an example, see Example 4.7.1. The choice of these methods is based on the likelihood of having a small standard deviation because the normalised measure will explode then. Using these two methods, we expect to have the lowest chance of an enormous normalised measure.

Example 4.7.1: Two methods to normalise our measures of interest

Assume we have three products with time frames:

$$\mathcal{T}_1 = \{\mathcal{T}_1^1, \mathcal{T}_1^{*1}, \mathcal{T}_1^2\}$$

$$\mathcal{T}_2 = \{\mathcal{T}_2^1, \mathcal{T}_2^{*1}, \mathcal{T}_2^2, \mathcal{T}_2^{*2}, \mathcal{T}_2^3\}$$

$$\mathcal{T}_3 = \{\mathcal{T}_3^1, \mathcal{T}_3^{*1}, \mathcal{T}_3^2\}$$

Products 1 and 3 have one OOS period, whereas product 2 has two OOS periods. We want to investigate residual $\epsilon_{i,t_{ja}}$ from Perspective 4.5.1.1. This means that we will have four different residuals and thus we will investigate the distribution of $(\epsilon_{1,t_{1a}}, \epsilon_{2,t_{1a}}, \epsilon_{2,t_{2a}}, \epsilon_{3,t_{1a}})$. However, we still need to normalise these residuals. We have two choices:

1. Divide all the residuals by $\hat{\sigma} = \hat{sd}(\epsilon_{1,t_{1a}}, \epsilon_{2,t_{1a}}, \epsilon_{2,t_{1a}}, \epsilon_{3,t_{1a}})$
2. Divide the residuals individually by $\hat{\sigma}_i = \hat{sd}((s_t)_{t\in\{\mathcal{T}_i^j,\dots,\mathcal{T}_i^{k_i+1}\}})$. Thus, $\hat{\sigma}_1 = \hat{sd}((s_t)_{t\in\{\mathcal{T}_1^1,\mathcal{T}_1^2\}})$, $\hat{\sigma}_2 = \hat{sd}((s_t)_{t\in\{\mathcal{T}_2^1,\mathcal{T}_2^2,\mathcal{T}_2^3\}})$, $\hat{\sigma}_3 = \hat{sd}((s_t)_{t\in\{\mathcal{T}_3^1,\mathcal{T}_3^2\}})$

To summarise, for every measure we investigate, we normalise every measure by the standard deviation of all the values for that measure, or we normalise each measure by the standard deviation of the product's in-stock sales data. To decide on these two choices, we have investigated them for all the measures.

In Figure 4.8 are densities of the two choices of normalisation for every measure. We have plotted the absolute values such that we can scale the x-axis into a log scale, note that as a result, the zero values are left out from the plot. The distance measure seems to explode, this is because certain distance values are high due to fitted $ARMA(1,1)$ processes that are almost non-stationary. Moreover, if we look at the densities, taking the standard deviation of a product's sales results in high outliers compared to the other method, except for measure $\beta_i^{j+1} - \beta_i^j$. These high values are unrealistic: a normalised residual of > 700 is not expected. Therefore we conclude that the normalisation by the standard deviation of the values of the measures is the best method to normalise. Even for the measure $\beta^{j+1} - \beta^j$, since these values stay in an acceptable range.

4.7.2. Results of the perspectives

First, we share an overview of how many products and OOS periods meet the assumptions given in Section 4.3 together with information about the post-processing steps we did to get the correct results. After that, we share the results of the perspectives for one OOS period (perspectives from Sections 4.4.1 and 4.5.4) followed by the results of the perspectives for multiple OOS periods (perspectives from Sections 4.4.2 and 4.5.5). Eventually, the results of the measures of interest are shared.

To investigate OOS periods, we have introduced assumptions in Section 4.3 that describe an OOS period. In summary, these assumptions are: an OOS period should have a length of two weeks min-

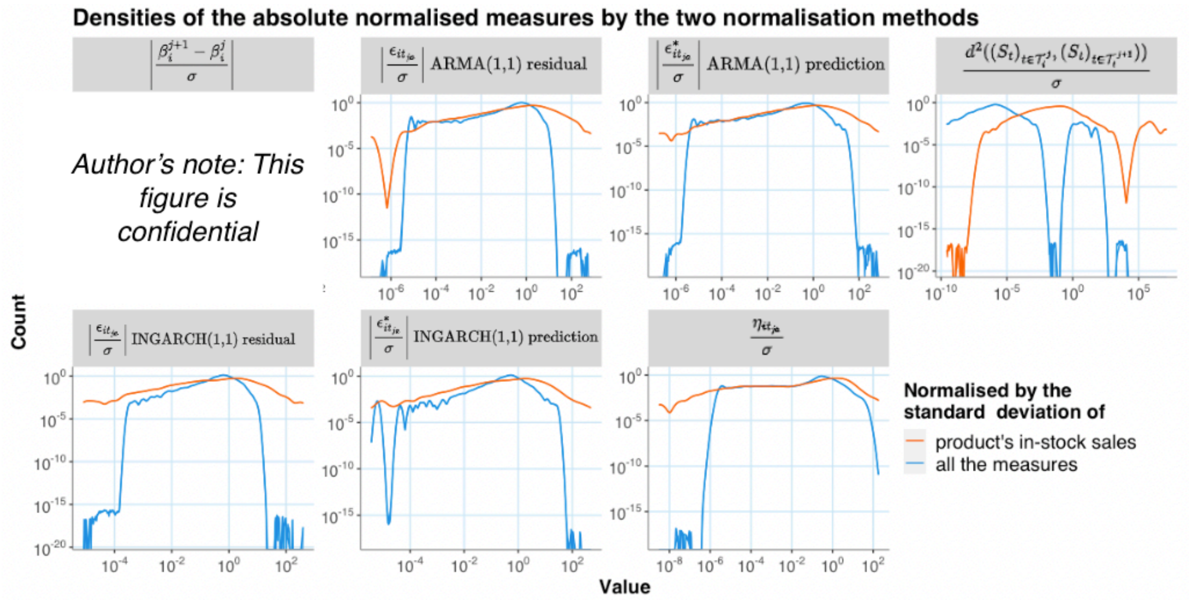

Figure 4.8: Densities of the absolute normalised measures by the two normalisation methods introduced in Example 4.7.1. Note that because of the log scales the zero values of the measures are left out of the plot.

imum and the consecutive in-stock days before and after the OOS period should have a length of one-week minimum. We have applied these assumptions to our data. Table 4.3 is the frequency table of the number of OOS periods that meet the assumptions. Approximately two-thirds of the product do not meet the assumptions, whereas one-third did (24,442 products). Most of these products have one OOS period that we can investigate. Thus, 19.40% of the products are being tested by the perspectives written for one OOS period, and 15.20% of the products are being tested by the perspectives written for multiple OOS periods. These are 13,705 and 10,737 products, respectively.

Overview OOS periods	
Percentage of the products	**Number of OOS periods**
65.40	0
19.40	1
8.55	2
3.88	3
1.74	4
0.69	5
0.24	6
0.09	7
0.01	8
2.83e-03	9

Table 4.3: Percentage of the products with the number of OOS periods that meet the assumptions introduced in Section 4.3.

We continue sharing the post-processing steps we do to interpret the results correctly. There are two post-processing steps that we do. The first step is for products that have constant sales before and after an OOS period, these are 17.84% of the 24,442 products. When we test the perspectives, a constant data input creates an error or false result for some of the perspectives. For example, the **t.test()** function that executes the t-test cannot handle constant input data. As a result, the p-value of the test does not exist. To solve this problem, we adjust the results manually because we can conclude that the OOS period does not influence sales for these products. As a result, we set all the p-values of the perspectives for these products equal to one.

The second post-processing step is for products with data for which $ARMA(1,1)$ or $INGARCH(1,1)$ processes cannot fit a stationary process. The fitted $ARMA(1,1)$ or $INGARCH(1,1)$ process should be stationary, else the distance metric in Definition 4.5.4 does not hold for example. As a result, the distance metric gets enormous. To avoid this problem, we analysed the fitted $ARMA(1,1)$ and $INGARCH(1,1)$ processes to detect non-stationary processes. For the non-stationary processes, we changed the results into *NA* (not available) values such that we will not use those results.

In the following sections, we will share the results of the perspectives. Note that when we say a perspective is being accepted/rejected, we mean that the null hypothesis of the hypothesis test in the perspective is being accepted/rejected.

Results of the perspectives for one OOS period Table 4.4 shows the results of the perspectives for one OOS period, in total 13,705 products with one OOS period are investigated. First, we remark that not every perspective could test all products. This problem only happens for time-dependent perspectives whenever the fitted $ARMA(1,1)$ or $INGARCH(1,1)$ process is not stationary. It becomes worse for Perspectives 4.5.2 and 4.5.4 because they depend on two separate process fits, increasing the chance of having a non-stationary fit.

Next, we see that the perspectives give surprisingly different results for the 5.01 % subset of the products: The smallest rejection percentage is 7.14 %, whereas the highest is 94.02 %. This wide range of rejection percentages indicates a lack of agreement between the perspectives. Thus, opposite conclusions can be drawn, which raises doubts about the meaning of these results. The different rejection percentages can be due to the different levels of complexity in the perspectives and the contradictions between some of them, for instance, the sales data for a given product cannot be both time-independent and dependent. Therefore, most perspectives are invalid, but we do not know which. Moreover, the perspectives' tests have different power, influencing rejection percentages. This is coherent with our discussion earlier in Section 4.4.1. Also, an insufficient sample size of the product's data can affect the performance of the tests.

Another remarkable result is that the rejection percentages between the 5.01 % subset of the products and all applicable products are not aligned. The difference is the largest at time-independent perspectives, we have a reason for this: These perspectives are applicable for all products, thus also for the products with constant sales. These products do not reject the perspectives, and therefore the rejection percentages will be lower. Moreover, these nonaligned percentages indicate instability in the perspectives: For Perspective 4.5.2 the subset size is approximately one-third of the applicable products set, which should indicate that the subset is a good representation. However, the rejection percentages differ almost by a factor of two.

Results of the perspectives for products with one OOS period

Perspective		Applicability perc.	Rejection perc. of applicable products	Rejection perc. of 5.01 % of the products
4.4.1	Difference in distribution	100.00	12.88	37.17
4.4.2	Larger in probability*	100.00	15.83	22.30
4.4.2	Smaller in probability**	100.00	15.61	27.70
4.4.3	Difference in means	100.00	25.14	43.59
4.4.4	Difference in Poisson param.	100.00	30.20	56.71
4.5.1.1	$ARMA(1,1)$ residual	46.90	10.72	8.61
4.5.1.2	$ARMA(1,1)$ prediction	32.32	12.46	11.37
4.5.2	$ARMA(1,1)$ distance metric	17.75	20.19	9.18
4.5.3.1	$INGARCH(1,1)$ residual	27.04	7.93	7.14
4.5.3.2	$INGARCH(1,1)$ prediction	25.62	9.77	9.62
4.5.4	Difference between $INGARCH(1,1)$ processes	12.13	92.06	94.02
4.5.5	$INGARCH(1,1)$ intervention effect	27.04	14.30	19.68

* This are the results of alternative hypothesis $H_{1.a}$.
** This are the results of alternative hypothesis $H_{1.b}$.

Table 4.4: The results of the perspectives for products with one OOS period. Only Perspective 4.4.5 is omitted since these results are investigated separately. The "Applicability percentage" is the percentage of products which the perspective could investigate. The "Rejection percentage of applicable products" is the percentage of applicable products that reject the hypothesis in the perspective. The "Rejection percentage of 5.01 % of the products" is another rejection percentage, but of the subset of products that all perspectives could investigate (except Perspective 4.4.5). These are 5.01 % of all products.

To further examine the difference in perspectives, we divide the products into subsets which reject the same time-independent or dependent perspectives. Figure 4.9 is the Venn diagram with the products that reject certain time-independent perspectives. The largest subset of products accepts all perspectives, and the second-largest subset rejects them all, this indicates that the perspectives are aligned. However, the second-largest subset rejects all perspectives except the difference in distribution perspective. This is remarkable since a difference in means should imply a difference in distributions, we already saw this remarkable result in our discussion about Perspectives 4.4.1, 4.4.2, 4.4.3 in Section 4.4.1. Apparently, Perspective 4.4.1 is unable to show this difference or less sensible to these differences. Moreover, the largest subset of products does not reject any perspective, a majority of the products' sales are not influenced by an OOS period according to the time-independent perspectives.

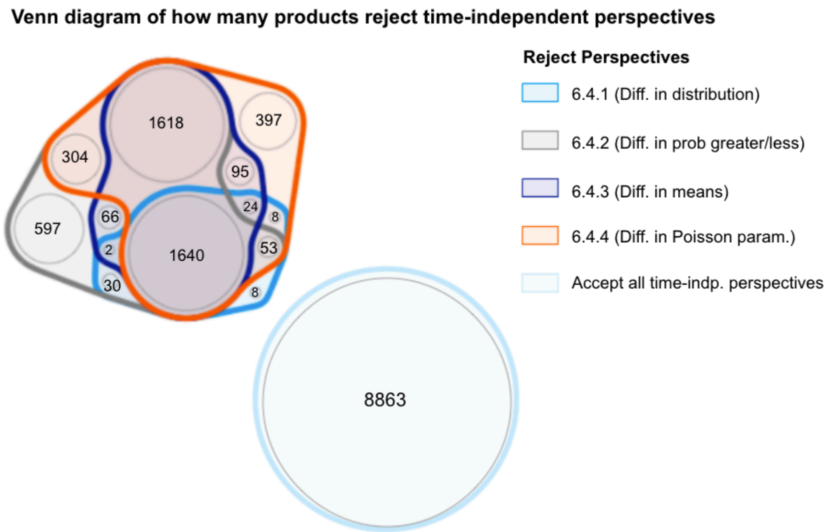

Figure 4.9: A Venn diagram of how many products reject time-independent perspectives. The products used in this Venn diagram are the products that could be tested for Perspectives 4.4.1, 4.4.2, 4.4.3 and 4.4.4. Note that we have combined the results of the perspective about larger or smaller in probability, since a product would not reject both these hypotheses.

Figure 4.10 is the Venn diagram for the time-dependent perspectives. The wide range of rejection percentages we saw in Table 4.4 is visible: one large subset of products only rejects Perspective 4.5.4. This influences the size of the subset of products that accept all perspectives. Also, there is a subset of products which rejects all the perspectives, however it is not the largest subset. The larger subsets only reject a few perspectives, again indicating the lack of agreement between the perspectives.

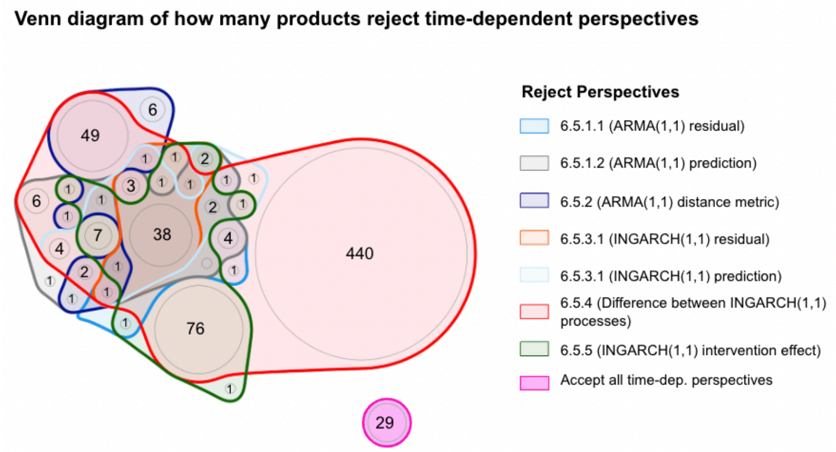

Figure 4.10: A Venn diagram of how many products reject time-dependent perspectives. The products used in this Venn diagram are the products that could be tested for Perspectives 4.5.1.1, 4.5.1.2, 4.5.2, 4.5.3.1, 4.5.3.2, 4.5.4 and 4.5.5.

We are left with analysing the results of Perspective 4.4.5, the perspective especially created for slow-moving products: Products with a maximum of one sale a day. In Table 4.5 we see a wide range of rejection percentages again. It seems like the more general perspectives cannot detect the differences in the datasets, whereas Perspective 4.4.5 does detect them. Thus the results of slow-moving products are dependent on the type of perspective.

Results of the time-independent perspectives of the slow-moving products		
Perspective		**Rejection perc.**
4.4.1	Difference in distribution	0.06
4.4.2	Larger in probability*	6.37
4.4.2	Smaller in probability**	3.27
4.4.3	Difference in means	4.95
4.4.4	Difference in Poisson param.	4.34
4.4.5	Difference in Bernoulli param.	22.34

* This are the results of alternative hypothesis $H_{1.a}$.
** This are the results of alternative hypothesis $H_{1.b}$.

Table 4.5: The results of the time-independent perspectives on the slow-moving products. These are 39.53 % of the products. The time-dependent perspectives are not considered since the used time series are not created especially for binary data.

Results of the perspectives for multiple OOS periods The perspectives for products with multiple OOS periods investigate 10,737 products with 29,194 OOS periods in total. Table 4.6 shows similar results as we saw in Table 4.4: Time-dependent perspectives are less applicable, and there is a wide range of rejection percentages. Thus, the perspectives for multiple OOS periods also lack agreement and have different powers. It is not unexpected that there is no difference in results between the perspectives for products with one OOS period and multiple OOS periods. Moreover, there are no results for a subset of the products for which all perspectives could test all OOS periods because this subset is too small to give representable results.

Furthermore, multiple alternative hypotheses are tested. Hypothesis $H_{1.a}$ is being rejected the most, followed by $H_{1.b}$ and $H_{1.c}$. This is expected given that the other two hypotheses are harder to reject. Remarkable are the high rejection percentages at the more complex perspectives for hypotheses $H_{1.b}$ and $H_{1.c}$ compared to other perspectives. This indicates that complex perspectives have more power.

Results of the perspectives for products with multiple OOS periods					
Perspective		**Applicability perc.**	**Rejection perc. of applicable products**		
			$H_{1.a}$	$H_{1.b}$	$H_{1.c}$
4.4.6	Difference in distribution	100.00	22.68	1.77	0.94
4.4.7	Larger in probability	100.00	25.26	0.68	0.82
4.4.7	Smaller in probability	100.00	30.08	0.84	1.14
4.4.8	Difference in means	100.00	38.10	4.55	3.13
4.4.9	Difference in Poisson param.	100.00	45.69	7.58	5.37
F.0.1.1	$ARMA(1,1)$ residual	72.45	14.97		3.53
F.0.1.2	$ARMA(1,1)$ prediction	58.98	19.86		5.59
F.0.2.1	$ARMA(1,1)$ distance metric	38.83	24.44	16.57	14.25
F.0.3.1	$INGARCH(1,1)$ residual	56.64	10.74		3.55
F.0.3.2	$INGARCH(1,1)$ prediction	52.81	15.82		3.92
F.0.4	Difference between $INGARCH(1,1)$ processes	31.69	92.39	84.40	80.66
F.0.5	$INGARCH(1,1)$ intervention effect	56.64	19.27		4.67

* This are the results of alternative hypothesis $H_{1.a}$.
** This are the results of alternative hypothesis $H_{1.b}$.

Table 4.6: The results of the perspectives for products with multiple OOS period. Only Perspective 4.4.10 is omitted. The ”Applicability percentage” is the percentage of products which the perspective could investigate. The ”Rejection percentage of applicable products” is the percentage of applicable products that reject the alternative hypothesis in the perspective, per alternative hypothesis. The ”Rejection percentage of 3.39 % of the products” is another rejection percentage, but of the subset of products that all perspectives could investigate (except Perspective 4.4.10). These are 3.39 % of all products. Some of the results of hypothesis $H_{1.b}$ miss because this hypothesis is not tested.

Results of the perspectives for multiple products We continue to the results of the perspective from Section 4.6, this perspective investigates the distribution of our measures of interest. First, we will share these perspectives' results. After that, we further investigate measure β_i^j.

All the products with an OOS period, one or multiple, are used for the results of this section, these are 42,889 OOS periods in total. Table 4.7 shows that almost all measures reject the tests if the mean or median equals zero. This indicates that OOS periods influence the sales. We investigate the measures further by analysing the box plots, see Figure 4.11. All the interquartile ranges are around zero, there are some outliers. The prediction measures have higher outliers than the residual ones, this indicates that the fitted processes are better at fitting than predicting. Moreover, the outliers of the difference in β_i^j measure are almost symmetrical.

Results of the measures of interest

Measure of interest θ^j	Applicability perc.	Reject $H_0: \mu(F) = 0$	Reject $H_0:$ med$(F) = 0$
β^j	100.00	No	Yes
$\epsilon_{t_{ja}}$ (residual $ARMA(1,1)$)	55.83	Yes	No
$\epsilon^*_{t_{ja}}$ (prediction $ARMA(1,1)$)	41.47	Yes	Yes
$d((s_t)_{t\in\mathcal{T}^j}, (s_t)_{t\in\mathcal{T}^{j+1}})^2$	22.54	Yes	Yes
$\epsilon_{t_{ja}}$ (residual $INGARCH(1,1)$)	37.25	Yes	Yes
$\epsilon^*_{t_{ja}}$ (prediction $INGARCH(1,1)$)	35.83	Yes	Yes
$\eta_{t_{ja}}$	37.25	Yes	Yes

Table 4.7: The results of Perspective 4.6.1 for every measure of interest. The "Applicability percentage" is the percentage of all the values of the measure which were available to use.

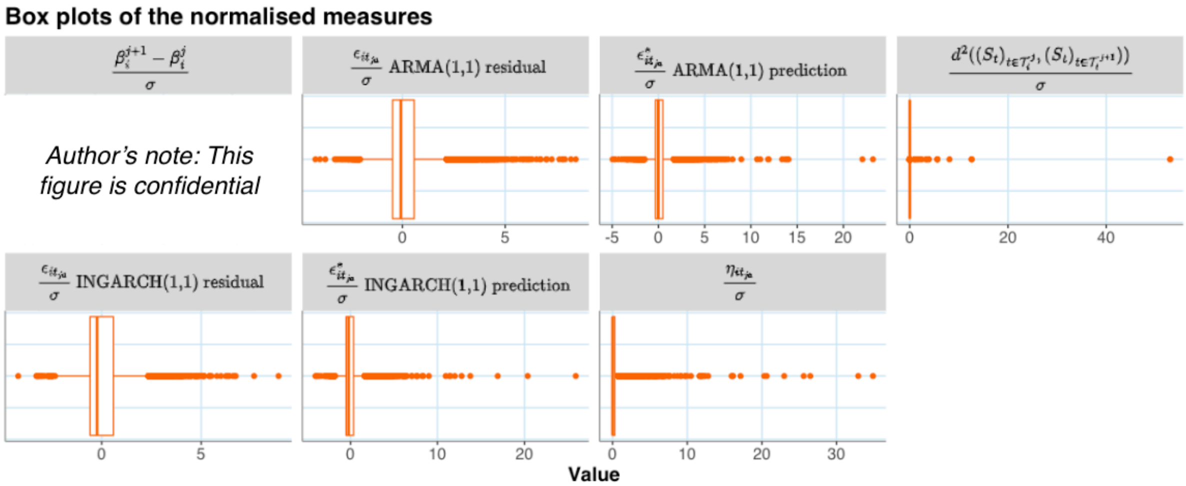

Figure 4.11: Box plots of the normalised measures of interest. The $ARMA(1,1)$ distance metric measure and $INGARCH(1,1)$ intervention effect measure only consist of positive values since the distance metric is squared and the intervention effect only can be positive by definition.

Further investigation of β_i^j We analyse the measure β_i^j to get more information about the influence of OOS periods on sales. Only β_i^j is being analysed because this measure is not dependent on a fitted process. First, we will look at the values of β_i^j and β_i^{j+1}. After that, we will look at specific characteristics of products and their values of β_i^j and β_i^{j+1}. We will name β_i^j and β_i^{j+1} the mean of the dataset before and after an OOS period.

The normalised means of the datasets before versus after an OOS period are plotted in Figure 4.12. Except for a few outliers, the scatter plot is symmetrical. Also, for small values of $\frac{\beta_i^j}{\sigma}$ and $\frac{\beta_i^{j+1}}{\sigma}$ diagonal lines appear, these are combinations of the means that differ a constant value. Moreover, we stated earlier that many sales data equal zero, this is also true for this investigation, see Table 4.8 for an overview of zero values in the means. We further investigate these zero means.

Author's note: This figure is confidential.

Figure 4.12: Scatter plot of the normalised means of the data before versus after an OOS period with identity line.

Author's note: This table is confidential.

Table 4.8: Overview of how many percentage of the normalised means equal zero or are larger than zero.

We have visualised the zero and non-zero normlaised means combinations of Table 4.8, see Figures 4.13a and 4.13b. Both density plots tell us that if one of the means is higher than five, the other mean is more likely to be non-zero, whereas if one of the means is close to zero, the other mean is less likely to be non-zero.

Author's note: This figure is confidential.

(a) Density of the normalised non-zero means of before an OOS period, divided over the zero and non-zero means after an OOS period.

Author's note: This figure is confidential.

(b) Density of the normalised non-zero means of after an OOS period, divided over the zero and non-zero means before an OOS period.

Figure 4.13: Density plots of the normalised non-zero means of the data before or after an OOS period .

We continue with the non-zero values of the normalised means, these are 73.81 % of all the means. To investigate the non-zero means better, we have estimated the two-dimensional kernel density, see Figure 4.14. Many data points are located around the identity line and symmetrically distributed. Thus, having a higher mean before an OOS period is almost as likely as having it after an OOS period. Moreover, 41.24 % of the points are under the identity line, 36.49 % are above, and 22.27 % are on the identity line: a slight majority of the average sales before an OOS period is higher than afterwards. This means that overall an OOS period slightly decreases the sales. Remark that these results are nil.

Author's note: This figure is confidential.

Figure 4.14: Two dimensional kernel density estimations of the non-zero values of $\frac{\beta_i^j}{\sigma}$ and $\frac{\beta_i^{j+1}}{\sigma}$ with the actuals plotted.

We further investigate OOS periods by looking at their specific characteristics. First, we look at which values of β_i^j and β_i^{j+1} reject Perspective 4.4.3, then we investigate if β_i^j and β_i^{j+1} behave differently per product types. Finally, we analyse if the means are related to the date the OOS period occurred.

Perspective 4.4.3 tests if the means of the sales before and after an OOS period are different. In Figure 4.15 we see that there are no points around the identity line that reject the perspective, this is logical. It is remarkable that some points that are not on the identity line accept the perspective, this means that the perspective fails to detect the different means. However, the points that are the most far away from the identity line do reject the perspective. These are the OOS periods with the highest disparity in average sales.

Author's note: This figure is confidential.

Figure 4.15: Two scatter plots of the non-zero values of $\frac{\beta_i^j}{\sigma}$ versus $\frac{\beta_i^{j+1}}{\sigma}$, the first plot are the values from OOS periods that accept Perspective 4.4.3, the second plot are values that reject this perspective.

Moreover, we analyse if the behaviour of the average sales is different for different types of products. To divide the products, we use a higher segment level than introduced in Section 2.4, namely product teams. There are 17 different product teams. Figure E.1 in Section E.3 shows the two-dimensional kernel density estimations of $\frac{\beta_i^j}{\sigma}$ and $\frac{\beta_i^{j+1}}{\sigma}$ per product team. Overall, they all have similarities around the identity lines. The dissimilarities are mainly at the lower levels of density. Furthermore, small shifts in the height of the means are seen. For example, the values of team Travel & Fitness and Cameras are lower than those of team Kitchen appliances big and Personal care. This means there is a difference in the level of sales between those teams.

We also calculate the correlation between $\frac{\beta_i^j}{\sigma}$ and $\frac{\beta_i^{j+1}}{\sigma}$ for every product team, see Table 4.9. All the correlations are positive but not very strong. Moreover, we investigate if the correlations differ significantly. To do this, we used the function **r.test()** of the **psych** [18] library. Figure E.2 in Section E.3 is a heat map of the test's p-values. Product teams Televisions & beamers and Garden & climate tools reject every test since their correlations are relatively low. Moreover, the majority of the tests are being rejected. We can conclude that the correlation between $\frac{\beta_i^j}{\sigma}$ and $\frac{\beta_i^{j+1}}{\sigma}$ of product teams differ significantly overall.

Correlation between $\frac{\beta_i^j}{\sigma}$ and $\frac{\beta_i^{j+1}}{\sigma}$ per product teams			
Product team	**Correlation**	**Product team**	**Correlation**
Audio	0.73	Personal care	0.66
Cameras	0.68	Travel & fitness	0.59
Gaming, printing & office	0.75	Smart home	0.73
Household	0.62	Phones, tablets & accessories	0.64
Kitchen appliances big	0.73	Televisions & beamers	0.53
Kitchen appliances small	0.66	Garden, tools & climate management	0.40
Coffee appliances	0.69	Washing machines & dryers	0.66
Laptops, desktops & accessoires	0.75	Wearables	0.74
Monitors, storage & components	0.70		

Table 4.9: Correlation between $\frac{\beta_i^j}{\sigma}$ and $\frac{\beta_i^{j+1}}{\sigma}$ for every product team.

Finally, we examine if the normalised means are related to the OOS periods' dates. In Figure 4.16 is the two dimensional kernel density estimation plotted of $\frac{\beta_i^j}{\sigma} - \frac{\beta_i^{j+1}}{\sigma}$ and the middle day of the OOS period (med($\mathcal{T}_i^{*j}$)). There are three periods for which a lot of OOS periods are located: Apr-Jun 2020, Dec-Feb 2020/21 and Aug-Oct 2021. Reasons for this could be Covid-19 lockdowns which result in increased sales of working-at-home supplies, Black-Friday shopping and Christmas shopping. Furthermore, the difference in normalised means has positive peaks during those same periods. Also, one negative peak is located around the beginning of 2021. We do not have a clarification for positive or negative peaks of $\frac{\beta_i^j}{\sigma} - \frac{\beta_i^{j+1}}{\sigma}$ at these time periods.

Author's note: This figure is confidential.

Figure 4.16: Two dimensional kernel density estimation of $\frac{\beta_i^j}{\sigma} - \frac{\beta_i^{j+1}}{\sigma}$ and the middle day of the OOS period (= med($\mathcal{T}_i^{*j}$)).

4.8. Conclusion

Multiple perspectives have investigated the influence of OOS periods by testing it differently. From the perspectives for individual products, we got results that are not aligned: the perspectives lack agreement or have different powers due to the perspectives' complexity and assumptions. As a result, the level of influence of an OOS period is dependent on the way it is tested. Thus we can conclude that there is some influence, but not how much.

Moreover, measures of interest were investigated. Their means and medians are not equal to zero, indicating that the OOS period influences the sales. However, their interquartile ranges show no extreme results that imply huge influences of OOS periods. Also, the means of the dataset before and after an OOS period were investigated deeper. The change in means after an OOS period is almost as likely to be positive as negative. Also, specific characteristics of the OOS periods or products do not influence the results. To conclude, the average sales after an OOS period differ but are as likely to increase as decrease.

5 Conclusion & Discussion

This thesis aims to improve Coolblue's direct demand estimation model for substitutable products. This is done by answering three research questions. With the first research question, we compare the three sub-models of the direct demand model to detect any weaknesses in them:

- *How do the three sub-models perform compared to each other?*

The sub-models have been evaluated on three criteria: applicability, best-performing and occurrence of large errors in the estimations. The linear model outperforms the other two models in its accuracy of the estimations, whereas the mean model is the most applicable to the products, but its estimations are poor. The EM model's estimations are also not outstanding, and the applicability of the sub-model is low.

Based on this conclusion, we want to improve the estimations of the mean model by answering the second research question:

- *Can we improve the estimated demand provided by the mean model?*

We introduced three new estimators that could replace the current estimator of the mean model under two criteria: The estimator should still be applicable to all products, and the average RRSE of the sales estimates should be lower. The new estimators use the product status or status-2 in their estimates. All three estimators meet the criteria, so the estimator with the lowest average RRSE is chosen as the new estimator. These estimates equal the average sales of a product on the days it has status available.

Finally, we did an extensive investigation about OOS periods to answer the last research question:

- *Do out-of-stock periods influence the sales? And if so, how can we react to this influence such that it does not affect the direct demand estimations.*

We created perspectives that investigate the influence of an OOS period in different ways. The perspectives indicate that the OOS periods influence sales, but these results are debatable: the rejection percentages were not aligned, indicating a lack of agreement between the perspectives, or different power of the tests or unstable perspectives. Thus, for some OOS periods, there is an influence on the sales. However, these results are dependent on the way it is tested. Furthermore, we further analysed the average sales before and after an OOS period. The OOS periods for which the means differ is the difference as likely to be positive as negative.

To answer the research question, a part of the products' OOS periods influence the sales, but it is hard to detect which products precisely have an influence since these results depend on how they are investigated. Therefore, no specific threshold or criteria can be set to mark an OOS period affecting sales. As a result, reacting to this influence is challenging.

5.1. Discussion and further research

The thesis advises Coolblue to improve its direct demand model, but also it gives the direction of thoughts and ideas for research. We share some of our thoughts and ideas for further research.

We improved the mean model by advising a new estimator with the lowest average RRSE over the estimated sales. However, the goal was to improve the direct demand estimations made by the mean model. Thus, we assume that the demand estimates are accurate if the sales are estimated accurately. This assumption can influence the mean model to become a sub-model that accurately estimates sales instead of demand. A solution could be only to evaluate the new estimators on a set of products that are a substitute for each other and are all in stock. As a result, there is no substitution, and there are no lost sales, thus the sales equal the direct demand. However, we do see the limitations of this solution: it is hard to find such a subset of products since 43.29 % of the time, products are not in stock, as we saw in Section 2.4.

Moreover, the OOS investigation can also be influenced by its assumptions. We consider multiple topics. First, we examine the sales to detect an influence of an OOS period. However, we do not consider any external influences, such as marketing campaigns, seasonality, competitors' product status, etc. All these factors can impact sales, and we do not observe this. Thus, a sudden increase or decrease can result from an external factor instead of an OOS period. For example, Kucuk (2004) [11] shows that in-store pricing influences a customer's decision. Thus, price differences in a product type can influence the sales. A topic for further research could be to create a model that also considers external factors when measuring the influence of an OOS period on sales.

Furthermore, we already mentioned it in our results of the OOS investigation. It is debatable that we investigate the products' sales data from different viewpoints: We assume that a product's sales data follows Poisson distribution and that it fits an $ARMA(1,1)$ and $INGARCH(1,1)$ process. However, these datasets cannot fit all these distributions/time series, thus a part of the fits has a low quality which results in invalid results. This problem is also visible in Table 4.5: Slow-moving products have different results per perspective. The perspective created for slow-moving products has a higher rejection percentage than the more generalised perspectives. To improve this, per product could be investigated which perspective best fits, such that the OOS periods are only investigated in optimal circumstances.

Continuing on this topic, we optimised the order of $ARMA(p,q)$ and $INGARCH(p,q)$ processes for all the products' sales data. However, these orders can also be chosen per product's sales data to optimise the fitted processes. This does take a lot of computation time.

Bibliography

[1] F. Bijma, M. Jonker, and A. Vaart. *An Introduction to Mathematical Statistics*. Amsterdam University Press, May 2017, pp. 105–159. DOI: 10.5117/9789462985100.

[2] J M. Bland and D. G Altman. “Statistics notes: Multiple significance tests: the Bonferroni method”. In: *BMJ* 310.6973 (Jan. 1995), pp. 170–170. DOI: 10.1136/bmj.310.6973.170.

[3] P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting*. New York, NY: Springer New York, 2002. DOI: 10.1007/b97391.

[4] W. J. Conover. *Practical Nonparametric Statistics.* 3rd. 4. John Wiley & Sons, Inc., Dec. 1999, pp. 456–462. ISBN: 978-0-471-16068-7.

[5] F.M. Dekking et al. *A Modern Introduction to Probability and Statistics: Understanding Why and How*. Delft: Springer, Mar. 2005, pp. 405–409. ISBN: 1852338962.

[6] M. Ezhil Kumar, D. P. Sharma, and A. V. Tapar. “Out-of-stock justifications and consumers’ behavioral outcomes– exploring the role of product type and sales level information in out-of-stock situations”. In: *Journal of Retailing and Consumer Services* 60 (May 2021). DOI: 10.1016/j.jretconser.2021.102458.

[7] M.P. Fay and M. A. Proschan. “Wilcoxon-Mann-Whitney or T-test? on assumptions for hypothesis tests and multiple interpretations of decision rules”. In: *Statistics Surveys* 4 (2010), pp. 1–39. DOI: 10.1214/09-SS051.

[8] J. Hadar and W. R. Russell. “Stochastic dominance and diversification”. In: *Journal of Economic Theory* 3.3 (Sept. 1971), pp. 288–305. DOI: 10.1016/0022-0531(71)90024-X.

[9] R. Helm, T. Hegenbart, and H. Endres. “Explaining costumer reactions to real stockouts”. In: *Review of Managerial Science* 7.3 (July 2013), pp. 223–246. DOI: 10.1007/s11846-012-0079-8.

[10] M. Kim and S. J. Lennon. “Consumer response to online apparel stockouts”. In: *Psychology and Marketing* 28.2 (Feb. 2011), pp. 115–144. DOI: 10.1002/mar.20383.

[11] S. U. Kucuk. “Reducing the out-of-stock costs in a developing retailing sector”. In: *Journal of International Consumer Marketing* 16.3 (Aug. 2004), pp. 75–104. DOI: 10.1300/J046v16n03{_}05.

[12] T. Liboschik, K. Fokianos, and R. Fried. “tscount : An R Package for Analysis of Count Time Series Following Generalized Linear Models”. In: *Journal of Statistical Software* 82.5 (2017), pp. 1–5. DOI: 10.18637/jss.v082.i05.

[13] E. A. Maharaj. “A significance test for classifying ARMA models”. In: *Journal of Statistical Computation and Simulation* 54.4 (1996), pp. 305–331. DOI: 10.1080/00949659608811737.

[14] M.D. Moran. “Arguments for rejecting the sequential Bonferroni in ecological studies”. In: *Oikos* 100.2 (Feb. 2003), pp. 403–405. DOI: 10.1034/j.1600-0706.2003.12010.x.

[15] P. J. Mumby. “Statistical power of non-parametric tests: A quick guide for designing sampling strategies”. In: *Marine Pollution Bulletin* 44.1 (2002), pp. 85–87. DOI: 10.1016/S0025-326X(01)00097-2.

[16] M. Murller. “Dynamic Time Warping”. In: *Information Retrieval for Music and Motion*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007. Chap. 4, pp. 69–84. DOI: 10.1007/978-3-540-74048-3{_}4.

[17] D. Piccolo. “A DISTANCE MEASURE FOR CLASSIFYING ARIMA MODELS”. In: *Journal of Time Series Analysis* 11.2 (Mar. 1990), pp. 153–164. DOI: 10.1111/j.1467-9892.1990.tb00048.x.

[18] W. Revelle. *Package 'psych': Procedures for Psychological, Psychometric, and Personality Research*. Tech. rep. 2022. URL: `https://personality-project.org/r/psych/`.

[19] J. A. Rice. *Mathematical statistics and data analysis*. Thomson/Brooks/Cole, 2007. ISBN: 0534399428.

[20] A. Sarda-Espinosa. *Package 'dtwclust': Time Series Clustering Along with Optimizations for the Dynamic Time Warping Distance*. Tech. rep. Feb. 2022. URL: `https://github.com/asardaes/dtwclust`.

[21] M. V. Shcherbakov et al. "A survey of forecast error measures". In: *World Applied Sciences Journal* 24.24 (2013), pp. 171–176. DOI: `10.5829/idosi.wasj.2013.24.itmies.80032`.

[22] C. R. C. Sheppard. "How Large should my Sample be? Some Quick Guides to Sample Size and the Power of Tests". In: *Marine Pollution Bulletin* 38.6 (June 1999), pp. 439–447. ISSN: 0025-326X.

[23] J. Son, J. H. Kang, and S. Jang. "The effects of out-of-stock, return, and cancellation amounts on the order amounts of an online retailer". In: *Journal of Retailing and Consumer Services* 51 (Nov. 2019), pp. 421–427. DOI: `10.1016/j.jretconser.2019.02.008`.

[24] S. I. Vrieze. "Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC)". In: *Psychological Methods* 17.2 (June 2012), pp. 228–243. DOI: `10.1037/a0027127`.

[25] U. Wadgave and M. Ravindra Khairnar. "Parametric test for non-normally distributed continuous data: For and against". In: *Electronic physician* 11.2 (Feb. 2019), pp. 7468–7470. DOI: `10.19082/7468`.

[26] F. Wilcoxon. "Individual Comparisons by Ranking Methods". In: *Biometrics Bulletin* 1.6 (Dec. 1945), p. 80. DOI: `10.2307/3001968`.

A

Unpaired two-samples Wilcoxon test results

The unpaired two-samples Wilcoxon test is used to compare to independent groups of samples [4]. It is an alternative to the unpaired two-samples t-test. For the t-test the data should be normally distributed. This is not the case with out data sets, therefore we use the Wilcoxon test.

A.1. Tests for Chapter 2

In tests 1 till 4, we want to know if $\mu_{s_{jt}}$ is significantly higher during Black Friday and Christmas for every country j. To test this, the week of Black Friday and the week before Christmas are labelled as set C. On the current dataset this set equals: $C = \{2020/12/21 - 2020/12/25, 2021/11/20 - 2021/11/26, 2021/12/20, 2021/12/21\}$. The set C' are all the dates from 2020/12/21 until 2021/12/20 that are not in C.

Test 1

- $H_0 : \mu_{s_{j,t\in C}} = \mu_{s_{j,t\in C'}} \quad \forall j$
- $H_1 : \mu_{s_{j,t\in C}} > \mu_{s_{j,t\in C'}} \quad \forall j$

Test 2

- $H_0 : \mu_{s_{NL,t\in C}} = \mu_{s_{NL,t\in C'}}$
- $H_1 : \mu_{s_{NL,t\in C}} > \mu_{s_{NL,t\in C'}}$

Test 3

- $H_0 : \mu_{s_{BE,t\in C}} = \mu_{s_{BE,t\in C'}}$
- $H_1 : \mu_{s_{BE,t\in C}} > \mu_{s_{BE,t\in C'}}$

Test 4

- $H_0 : \mu_{s_{GER,t\in C}} = \mu_{s_{GER,t\in C'}}$
- $H_1 : \mu_{s_{GER,t\in C}} > \mu_{s_{GER,t\in C'}}$

From Table A.1 we see that all tests reject the null hypothesis, except the test for Germany. Thus the sales are higher during the Black Friday and Christmas days in all countries except Germany.

In tests 5 we want to test if the sales of products for which $\mu_{cc_{ij}} \geq 50\%$ are higher than for products for which $\mu_{cc_{ij}} < 50\%$.

Results tests	
Test	**P-value**
1	5.41e-06
2	1.04e-05
3	<1.44e-06
4	<0.24

Table A.1: Results of the significance tests 1 till 4.

Test 5

- $H_0 : \{s_{ij} | \forall i,j,\ \mu_{cc_{ij}} \geq 50\%\} = \{s_{ij} | \forall i,j : \mu_{cc_{ij}} < 50\%\}$
- $H_1 : \{s_{ij} | \forall i,j,\ \mu_{cc_{ij}} \geq 50\%\} > \{s_{ij} | \forall i,j : \mu_{cc_{ij}} < 50\%\}$

Results tests	
Test	**P-value**
5	<2.2e-16

Table A.2: Results of the significance test 5.

The test rejects the null hypothesis. Thus the sales of products for which $\mu_{ccij} \geq 50\%$ are significantly higher than products for which $\mu_{ccij} < 50\%$

A.2. Tests for Chapter 3

In tests 6,7 and 8 we test if the highest 25 % of the RRSEs of every sub-model are significantly different. Let ϵ_{ms} be the RRSEs of all the sales estimations made by sub-model $ms \in \{Mean, Linear, EM\}$ and let $Q_3(\epsilon_{ms})$ be all the values above third quartile of all the RRSE values for one model ms (highest 25 % of the values).

Test 6

- $H_0 : Q_3(\epsilon_{EM}) = Q_3(\epsilon_{Linear})$
- $H_1 : Q_3(\epsilon_{EM}) > Q_3(\epsilon_{Linear})$

Test 7

- $H_0 : Q_3(\epsilon_{Mean}) = Q_3(\epsilon_{Linear})$
- $H_1 : Q_3(\epsilon_{Mean}) > Q_3(\epsilon_{Linear})$

Test 8

- $H_0 : Q_3(\epsilon_{Mean}) = Q_3(\epsilon_{EM})$
- $H_1 : Q_3(\epsilon_{Mean}) > Q_3(\epsilon_{EM})$

Results test	
Test	**P-value**
6	<2.2e-16
7	<2.2e-16
8	<2.23e-09

Table A.3: Results of the significance tests 6 till 8.

All tests reject the null hypothesises. Thus the highest 25 % RRSEs of every sub-model are significantly different.

B

Detailed explanation of the EM model

Author's note: This chapter is confidential.

C

Performance of the new estimators for the mean model per cluster

We want to investigate if the estimators introduced in Section 3.3.1 performed differently on different types of products. Therefore we cluster the products based on their sales.

The **tsclust()** function from the **dtwclust** [20] library is used, it is a library specialised in time series clustering using Dynamic Time Warping (DTW) [16]. DTW is a technique to measure similarity between two time series that do not align exactly in time. As distance input in the function we chose ”dtw_basic” and as centroid input we chose ”mean”.

The number of clusters, k, is optimised by evaluating the quality of the cluster. For 5500 randomly picked products, we compare the results for $k = 2, ..., 10$ using the Silhouette index. It measures the distance between clusters, ranging between -1 and 1, a high value indicates a good cluster size. Note that before the clustering, we standardise the data. In Figure C.1 are the Silhouette indexes plotted against k. For $k = 4$ the index has an optimum, therefore we will make four clusters.

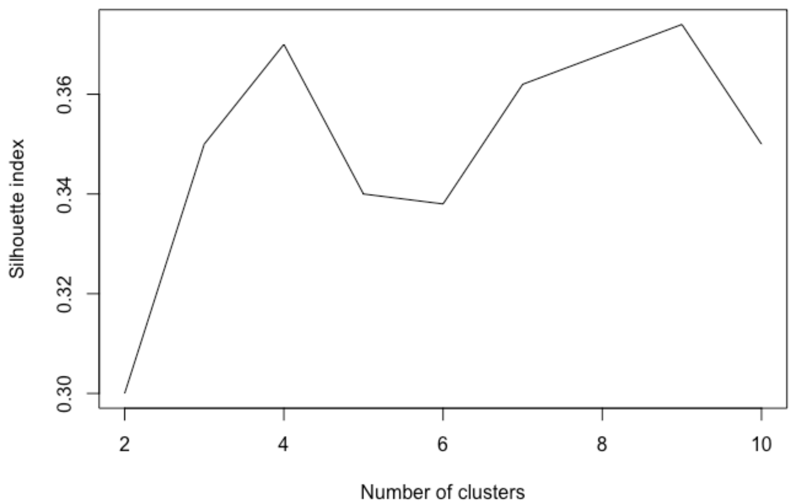

Figure C.1: The silhouette index for every cluster size.

Let $\mu_{NLt} = \frac{1}{n_{NL}} \sum_{i=1}^{n_{NL}} \mu_{iNLt}$ with n_{NL} = # products in NL, be the average daily sales of the products that we are investigating. Moreover, let μ_{NLt}^{j} be the average daily sales of the products in cluster $j = 1, ..., 4$.

In Figure C.2 is μ_{NLt}^{j} plotted of every cluster j. We can see the difference in sales of the clusters. We are especially interested in the difference in performance between clusters 1 & 2 and clusters 3 & 4, because the difference are large between those groups of clusters.

Author's note: This figure is confidential.

Figure C.2: The average daily sales of the four clusters plotted as time series.

After investigating the results of the estimators per cluster, we concluded that dividing the data into clusters does not influence the performance of the estimators. In every cluster, the same estimator came out as the best estimator.

D

Derivations of equations

D.1. Test statistic TS_{1,T^a} for comparing two Poisson distributions

To test Perspective 4.4.4 we create a test statistic using the asymptotic distribution of the MLE: For the MLE estimator $\hat{\lambda}$ with real value λ the following holds [19]:

$$\sqrt{TI(\hat{\lambda})}(\hat{\lambda} - \lambda) \xrightarrow{d} N(0, 1), \tag{D.1}$$

with $I(\cdot)$ the Fisher information and T the sample size. If we assume the null hypothesis of Perspective 4.4.4: $\lambda_1 = \lambda_2$, we can write Equation D.1 for λ_1 and λ_2 and combine them:

$$\hat{\lambda}_1 \xrightarrow{d} N(\lambda_1, (T^a I(\hat{\lambda}_1))^{-1}), \qquad \hat{\lambda}_2 \xrightarrow{d} N(\lambda_2, (T^a I(\hat{\lambda}_2))^{-1}), \tag{D.2}$$

$$(\hat{\lambda}_2 - \hat{\lambda}_1) \xrightarrow{d} N(\lambda_2 - \lambda_1, (T^a I(\hat{\lambda}_1))^{-1} + (T^a I(\hat{\lambda}_2))^{-1}) = N(0, 2(T^a I(\hat{\lambda}_a))^{-1}), \tag{D.3}$$

$$\sqrt{\frac{T^a}{2} I(\hat{\lambda}_a)}(\hat{\lambda}_2 - \hat{\lambda}_1) \xrightarrow{d} N(0, 1), \tag{D.4}$$

with $\hat{\lambda}_a = \frac{\hat{\lambda}_1 + \hat{\lambda}_2}{2}$, since we are trying to approach the values of $\hat{\lambda}_1$ and $\hat{\lambda}_2$. The Fisher information for the Poisson distribution equals $\frac{1}{\lambda}$. Therefore, we test if $\lambda_1 = \lambda_2$ with test statistic $TS_{1,T^a} = \sqrt{\frac{T^a}{2\hat{\lambda}_a}}(\hat{\lambda}_2 - \hat{\lambda}_1)$ and the two-sided Z-score table.

D.2. Distance measure for $ARMA(1, 1)$

Let X_t be an $ARMA(1, 1)$ process:

$$X_t = \phi X_{t-1} + Z_t + \theta Z_{t-1}, \qquad Z_t \sim WN(0, \sigma^2). \tag{D.5}$$

We can rewrite this into:

$$\phi(B) X_t = \theta(B) Z_t, \tag{D.6}$$

with,

$$\phi(z) = 1 - \phi z \qquad \theta(z) = 1 + \theta z. \tag{D.7}$$

Then,

$$\pi(z) = \frac{\phi(z)}{\theta(z)} = \frac{1 - \phi z}{1 + \theta z}, \tag{D.8}$$

$$(1 + \theta z)(1 + \pi_1 z + \pi_2 z^2 + \ldots) = 1 - \phi z. \tag{D.9}$$

This results in:

$$\pi_i = -(\phi + \theta)(-\theta)^{i-1},\ i > 0. \tag{D.10}$$

Using Equation D.10 for the distance metric in Definition 4.5.4 for $ARMA(1,1)$ processes X_t and Y_t gives:

$$d(X_t, Y_t) = \sqrt{\sum_{j=1}^{\infty}(-(\phi_x + \theta_x)(-\theta_x)^{j-1} + (\phi_y + \theta_y)(-\theta_y)^{j-1})^2}. \tag{D.11}$$

Then using the geometric series $\sum_{j=1}^{\infty} a^j = \frac{a}{1-a}$ for $|a| < 1$, the distance metric becomes:

$$\begin{aligned} d(X_t, Y_t) &= \sqrt{\sum_{j=1}^{\infty}((\phi_x + \theta_x)^2(-\theta_x)^{2j-2} + (\phi_y + \theta_y)^2(-\theta_y)^{2j-2} - 2(\phi_x + \theta_x)(\phi_y + \theta_y)(-\theta_x)^{j-1}(-\theta_y)^{j-1})} \\ &= \sqrt{\frac{(\phi_x + \theta_x)^2}{\theta_x^2}\frac{\theta_x^2}{1 - \theta_x^2} + \frac{(\phi_y + \theta_y)^2}{\theta_y^2}\frac{\theta_y^2}{1 - \theta_y^2} - 2\frac{(\phi_x + \theta_x)(\phi_y + \theta_y)}{\theta_x\theta_y}\frac{\theta_x\theta_y}{1 - \theta_x\theta_y}} \\ &= \sqrt{\frac{(\phi_x + \theta_x)^2}{1 - \theta_x^2} + \frac{(\phi_y + \theta_y)^2}{1 - \theta_y^2} - 2\frac{(\phi_x + \theta_x)(\phi_y + \theta_y)}{1 - \theta_x\theta_y}}. \end{aligned} \tag{D.12}$$

D.3. Test statistic TS_{4,T^a} for comparing two $ARMA(1,1)$ processes

We will write out some of the elements in Equation 4.19, the test statistic to compare two $ARMA(1,1)$ processes.. Let X_t and Y_t be two stationary $ARMA(1,1)$ processes:

$$X_t = \phi_x X_{t-1} + Z_t + \theta_x Z_{t-1} \quad Z_t \sim WN(0, \sigma_x^2), \tag{D.13}$$

$$Y_t = \phi_y Y_{t-1} + V_t + \theta_y V_{t-1} \quad V_t \sim WN(0, \sigma_y^2). \tag{D.14}$$

First of all, in Section D.2 we already saw the computation of the π-values, see Equation D.10. Thus vectors $\hat{\boldsymbol{\pi}}_x$ and $\hat{\boldsymbol{\pi}}_y$ become:

$$\hat{\boldsymbol{\pi}}_x = (-\hat{\phi}_x - \hat{\theta}_x, -\hat{\theta}_x(-\hat{\phi}_x - \hat{\theta}_x))^T, \qquad \hat{\boldsymbol{\pi}}_y = (-\hat{\phi}_y - \hat{\theta}_y, -\hat{\theta}_y(-\hat{\phi}_y - \hat{\theta}_y))^T. \tag{D.15}$$

Matrices $\hat{R}_x(k)$ and $\hat{R}_y(k)$ are the upper $k \times k$ sub-matrices of the estimated infinite auto-correlation matrices. For an $ARMA(1,1)$ process k equals two, thus the matrices become:

$$\hat{R}_x(2) = \begin{pmatrix} 1 & \hat{\rho}_x(1) \\ 0 & 1 \end{pmatrix}, \qquad \hat{R}_y(2) = \begin{pmatrix} 1 & \hat{\rho}_y(1) \\ 0 & 1 \end{pmatrix}. \tag{D.16}$$

For an $ARMA(1,1)$ process the auto-correlation equals:

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{\sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+1}}{\sigma^2 \sum_{j=0}^{\infty} \psi_j^2}, \tag{D.17}$$

where ψ_j-values come from the coefficients of the process:

$$\psi(z) = \frac{\theta(z)}{\phi(z)}, \tag{D.18}$$

for an $ARMA(1,1)$ these values equal:

$$\psi_i = (\phi + \theta)\psi^{i-1}, \qquad i > 0. \tag{D.19}$$

Thus, $\rho(1)$ becomes:

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{\sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+1}}{\sigma^2 \sum_{j=0}^{\infty} \psi_j^2} = \frac{\theta + \phi + \frac{(\theta+\phi)^2\phi}{1-\phi^2}}{1 + \frac{(\theta+\phi)^2}{1-\phi^2}}. \tag{D.20}$$

Finally, the matrix $\hat{V}$ becomes:

$$\begin{aligned} \hat{V} = (\hat{\sigma}_x^2 \hat{R}_x^{-1}(2) + \hat{\sigma}_y^2 R_y^{-1}(2)) &= \hat{\sigma}_x^2 \begin{pmatrix} 1 & -\hat{\rho}_x(1) \\ 0 & 1 \end{pmatrix} + \hat{\sigma}_y^2 \begin{pmatrix} 1 & -\hat{\rho}_y(1) \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \hat{\sigma}_x^2 + \hat{\sigma}_y^2 & -\hat{\sigma}_x^2 \hat{\rho}_x(1) - \hat{\sigma}_y^2 \hat{\rho}_y(1) \\ 0 & \hat{\sigma}_x^2 + \hat{\sigma}_y^2 \end{pmatrix}, \end{aligned} \tag{D.21}$$

and the test statistic equals:

$$TS_{4,T^a} = T^a (\hat{\boldsymbol{\pi}}_x - \hat{\boldsymbol{\pi}}_y)^T \begin{pmatrix} \hat{\sigma}_x^2 + \hat{\sigma}_y^2 & -\hat{\sigma}_x^2 \hat{\rho}_x(1) - \hat{\sigma}_y^2 \hat{\rho}_y(1) \\ 0 & \hat{\sigma}_x^2 + \hat{\sigma}_y^2 \end{pmatrix}^{-1} (\hat{\boldsymbol{\pi}}_x - \hat{\boldsymbol{\pi}}_y). \tag{D.22}$$

D.4. Test statistic TS_{5,T^a} for comparing two $INGARCH(1,1)$ processes

To compute a test statistic for comparing the coefficients of two $INGARCH(1,1)$ processes we use the asymptotical normality of the coefficients given by Liboschik et al. (2017) [12]. Let Y_t be a stationary $INGARCH(1,1)$ process:

$$X_t | \mathcal{F}_{t-1} \sim Pois(\lambda_t) \qquad \lambda_t = \alpha + \phi X_{t-1} + \theta \lambda_{t-1}, \tag{D.23}$$

with $\mathcal{F}_t = \sigma(X_i,\ i \leq t)$ the history of the process X_t and $\alpha, \phi, \theta > 0$ constants. For stationarity $\phi + \theta < 1$ must hold. Define $\boldsymbol{\omega} = (\alpha, \phi, \theta)^T$ the vector of the parameters. Then, for vector $\boldsymbol{\omega}$ the following holds:

$$\sqrt{T}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}) \xrightarrow{d} N_3(0, G_T^{-1}(\hat{\boldsymbol{\omega}})), \tag{D.24}$$

as $T^a \to \infty$, with T the sample size, $\hat{\boldsymbol{\omega}}$ the estimator of $\boldsymbol{\omega}$ and $G_T(\cdot)$ the conditional information matrix. The conditional information matrix equals:

$$G_T(\boldsymbol{\omega}) = \sum_{t=1}^{T} \left(\frac{1}{\lambda_t(\boldsymbol{\omega})} \right) \left(\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} \right) \left(\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} \right)^T. \tag{D.25}$$

First we will work out the test statistic, then we will work out matrix G. Assume we have two stationary $INGARCH(1,1)$ processes, X_t and Y_t:

$$X_t|\mathcal{F}_{1,t-1} \sim Pois(\lambda_{1,t}) \qquad \lambda_{1,t} = \alpha_x + \phi_x X_{t-1} + \theta_x \lambda_{1,t-1}, \qquad \boldsymbol{\omega}_x = (\alpha_x, \phi_x, \theta_x)^T, \tag{D.26}$$

$$Y_t|\mathcal{F}_{2,t-1} \sim Pois(\lambda_{2,t}) \qquad \lambda_{2,t} = \alpha_y + \phi_y Y_{t-1} + \theta_y \lambda_{2,t-1}, \qquad \boldsymbol{\omega}_y = (\alpha_y, \phi_y, \theta_y)^T. \tag{D.27}$$

Then,

$$\sqrt{T^x}(\hat{\boldsymbol{\omega}}_x - \boldsymbol{\omega}_x) \xrightarrow{d} N_3(0, G_{T^x}^{-1}(\hat{\boldsymbol{\omega}}_x)), \qquad \sqrt{T^y}(\hat{\boldsymbol{\omega}}_y - \boldsymbol{\omega}_y) \xrightarrow{d} N_3(0, G_{T^y}^{-1}(\hat{\boldsymbol{\omega}}_y)). \tag{D.28}$$

We assume $\boldsymbol{\omega}_x = \boldsymbol{\omega}_y$, using this we can rewrite the asymptotic normalities into:

$$(\hat{\boldsymbol{\omega}}_x - \hat{\boldsymbol{\omega}}_y) \xrightarrow{d} N_3(\boldsymbol{\omega}_x - \boldsymbol{\omega}_y, (T^x)^{-1}G_{T^x}^{-1}(\hat{\boldsymbol{\omega}}_x) + (T^y)^{-1}G_{T^y}^{-1}(\hat{\boldsymbol{\omega}}_y)) = N_3(0, 2(T^a)^{-1}G_{T^a}^{-1}(\hat{\boldsymbol{\omega}}_a)), \tag{D.29}$$

$$\sqrt{\frac{T^a G_{T^a}(\hat{\boldsymbol{\omega}}_a)}{2}}(\hat{\boldsymbol{\omega}}_x - \hat{\boldsymbol{\omega}}_y) \xrightarrow{d} N_3(0, I_3), \tag{D.30}$$

with $T^a = T^x = T^y$ and $\hat{\boldsymbol{\omega}}_a = 1/2\hat{\boldsymbol{\omega}}_x + 1/2\hat{\boldsymbol{\omega}}_y$. Let $X = \sqrt{\frac{T^a G_{T^a}(\hat{\boldsymbol{\omega}}_a)}{2}}(\hat{\boldsymbol{\omega}}_x - \hat{\boldsymbol{\omega}}_y)$, then $X \sim N_3(0, I_3)$ with I_3 a positive definite matrix. We can rewrite this into:

$$(X-0)^T I_3^{-1}(X-0) \sim \chi_3^2, \tag{D.31}$$

$$X^T X = ||X||_2^2 = \left\| \sqrt{\frac{T^a G_{T^a}(\hat{\boldsymbol{\omega}}_a)}{2}}(\hat{\boldsymbol{\omega}}_x - \hat{\boldsymbol{\omega}}_y) \right\|_2^2 \sim \chi_3^2. \tag{D.32}$$

Therefore, we can test if the coefficients of two $INGARCH(1,1)$ process are the same with test statistic TS_{5,T^a} and the chi-square distribution table.

$$TS_{5,T^a} = \left\| \sqrt{\frac{T^a G_{T^a}(\hat{\boldsymbol{\omega}}_a)}{2}}(\hat{\boldsymbol{\omega}}_x - \hat{\boldsymbol{\omega}}_y) \right\|_2^2. \tag{D.33}$$

Now we are only left to write out matrix $G_T(\boldsymbol{\omega})$, see Equation D.25. We will start by writing out $\left(\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\right)^T$ for $INGARCH(1,1)$ process described in Equation D.23 :

$$\left(\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \boldsymbol{\omega}}\right)^T = \left(\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \alpha}, \frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \phi}, \frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \theta}\right), \tag{D.34}$$

$$\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \alpha} = 1 + \theta \frac{\partial \lambda_{t-1}(\boldsymbol{\omega})}{\partial \alpha}, \tag{D.35}$$

$$\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \phi} = Y_{t-1} + \theta \frac{\partial \lambda_{t-1}(\boldsymbol{\omega})}{\partial \phi}, \tag{D.36}$$

$$\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \theta} = \lambda_{t-1} + \theta \frac{\partial \lambda_{t-1}(\boldsymbol{\omega})}{\partial \theta}. \tag{D.37}$$

In Equations D.35, D.36 and D.37, there is a recursion of the derivative of $\lambda_t(\boldsymbol{\omega})$. To solve this recursion, we use a method provided by Liboschik et al. (2017) [12], which is to initialise λ_t by the respective marginal expectations. This means:

$$\lambda_0 = E(Y_t) = E(\lambda_t) = \frac{\alpha}{1-\phi-\theta}. \tag{D.38}$$

This results in:

$$\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \alpha} = \sum_{j=0}^{t-1} \theta^j + \frac{\theta^t}{1-\phi-\theta}, \tag{D.39}$$

$$\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \phi} = \sum_{j=1}^{t} Y_{t-j}\theta^{j-1} + \frac{\alpha\theta^t}{(1-\phi-\theta)^2}, \tag{D.40}$$

$$\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \theta} = \sum_{j=1}^{t} \lambda_{t-j}\theta^{j-1} + \frac{\alpha\theta^t}{(1-\phi-\theta)^2}, \tag{D.41}$$

Thus, matrix $G_T(\boldsymbol{\omega})$ becomes:

$$G_T(\boldsymbol{\omega}) = \sum_{t=1}^{T} \left(\frac{1}{\lambda_t(\boldsymbol{\omega})}\right) \begin{pmatrix} \frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \alpha}\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \alpha} & \frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \alpha}\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \phi} & \frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \alpha}\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \theta} \\ \frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \phi}\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \alpha} & \frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \phi}\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \phi} & \frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \phi}\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \theta} \\ \frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \theta}\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \alpha} & \frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \theta}\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \phi} & \frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \theta}\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \theta} \end{pmatrix}, \tag{D.42}$$

with Equations D.39, D.40 and D.41 for $\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \alpha}$, $\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \phi}$ and $\frac{\partial \lambda_t(\boldsymbol{\omega})}{\partial \theta}$, respectively.

E

Additional figures and tables

A part of the figures and tables are in this chapter to reduce space in the main matter of the thesis.

E.1. Overview of failures of the arima() function

In Table E.1 are the percentages of the products given for which the **arima()** function failed to fit an $ARIMA(p,q)$ model. There are multiple reasons for the failure, we have decided to ignore those. In Table E.1 we see that the percentage of failures increases with the values of p and q. The higher the order of the model becomes, the more failures occur.

Percentage of the products for which function arima() failed per *p* and *q*

p \ *q*	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
0	0.00	0.13	0.43	0.77	1.20	1.47	1.76	2.23	2.69	3.23	3.31	3.94	4.49	5.04	5.47	6.33
1	0.04	2.80	4.06	3.79	4.07	4.60	4.54	5.21	5.61	6.00	6.97	7.79	8.23	8.94	9.36	9.90
2	0.06	4.69	8.11	8.83	9.23	9.27	9.34	9.50	10.79	11.83	12.77	13.60	14.67	15.19	16.09	16.87
3	0.14	3.33	9.27	14.99	15.13	16.29	16.64	17.64	17.86	19.06	20.86	21.46	23.24	24.69	25.31	24.90
4	0.24	3.57	9.54	17.14	23.17	23.83	24.34	25.40	25.97	27.54	27.81	29.51	30.41	31.86	31.69	33.26
5	0.31	3.61	10.20	18.03	24.47	31.11	31.97	32.79	32.61	33.39	35.51	36.43	37.40	38.41	38.44	38.83
6	0.37	3.59	10.27	18.53	26.07	32.34	37.13	37.66	37.90	39.73	41.26	42.44	41.16	43.73	43.31	44.16
7	0.46	3.76	10.54	18.10	25.96	33.64	38.51	42.46	43.31	44.61	45.47	45.40	46.96	47.43	47.19	47.41
8	0.63	3.73	11.50	18.96	27.46	34.19	39.19	44.19	47.26	48.84	48.60	49.37	50.07	49.67	50.94	51.31
9	0.80	4.27	11.23	20.01	27.60	34.30	40.81	44.93	49.33	51.36	51.91	51.80	52.29	52.46	54.23	54.91
10	0.94	4.10	12.54	20.89	27.69	36.20	41.04	45.90	49.30	52.73	54.94	55.13	56.06	55.54	56.84	57.87
11	1.06	4.46	12.73	21.10	30.13	37.30	42.10	46.44	49.54	53.30	56.14	58.60	58.44	59.54	59.86	60.47
12	1.34	5.13	13.36	22.21	30.34	38.36	43.19	47.10	50.11	53.59	57.49	59.53	61.46	61.99	61.39	62.76
13	1.46	5.09	13.71	23.00	31.50	37.76	42.76	47.89	50.16	54.54	58.00	60.21	62.37	65.16	64.93	66.34
14	1.76	5.26	13.77	23.43	32.10	37.59	43.60	47.73	50.41	54.06	56.94	60.17	62.43	65.30	68.57	67.34
15	1.93	5.76	13.94	23.14	32.54	39.81	43.71	47.43	50.91	54.49	57.74	61.33	63.44	66.51	68.66	71.30

Table E.1: Percentage of the products per p and q value for which the function **arima()** failed to fit an $ARMA(p,q)$ process due to different reasons.

E.2. Summaries of all the perspectives

Tables E.2 and E.3 are an overview of all the perspectives with the tests that are used to test them.

Summary of all the time-independent perspectives			
Perspective		**Alt. hypothesis**	**Test**
4.4.1	Diff. in distributions	1	K-S test
4.4.2	Diff. in probability	1.a, 1.b	WMW test
4.4.3	Diff. in means	1	T-test
4.4.4	Diff. in Poisson param.	1	Test statistic TS_{1,T^a} with two-sided Z-score table
4.4.5	Diff. in Bernoulli param.	1	Binomial test
4.4.6	Diff. in distributions	1.a, 1.b	Multiple K-S test with Bonferroni correction
4.4.6	Diff. in distributions	1.c	Multiple K-S test without Bonferroni correction
4.4.7	Diff. in probability	1.a.a, 1.a.b, 1.b.a, 1.b.b	Multiple WMW test with Bonferroni correction
4.4.7	Diff. in probability	1.c.a, 1.c.b	Multiple WMW test without Bonferroni correction
4.4.8	Diff. in means	1.a, 1.b	Multiple t-test with Bonferroni correction
4.4.8	Diff. in means	1.c	Multiple t-test without Bonferroni correction
4.4.9	Diff. in Poisson param.	1.a, 1.b	Test statistic $TS_{1,T^{i,j,a},i,j}$ with two-sided Z-score table with Bonferroni correction
4.4.9	Diff. in Poisson param.	1.c	Test statistic $TS_{1,T^{i,j,a},i,j}$ with two-sided Z-score table without Bonferroni correction
4.4.10	Diff. in Bernoulli param.	1.a, 1.b	Multiple Binomial test with Bonferroni correction
4.4.10	Diff. in Bernoulli param.	1.c	Multiple Binomial test without Bonferroni correction

$TS_{1,T^a} = \sqrt{\frac{T^a}{2\hat{\lambda}_a}}(\hat{\lambda}_2 - \hat{\lambda}_1)$.

$TS_{1,T^{i,j,a},i,j} = \sqrt{\frac{T^{i,j,a}}{2\hat{\lambda}_{i,j,a}}}(\hat{\lambda}^j - \hat{\lambda}^i)$.

Table E.2: A summary of all the time-independent perspectives: A short description of the perspective with the tests per alternative hypothesis.

Summary of all the time-dependent perspectives

Perspective		Alt. hypothesis	Test
4.5.1.1	$ARMA(1,1)$ residual	1	Test statistic TS_2 with two-sided Z-score table
4.5.1.2	$ARMA(1,1)$ prediction	1	Test statistic TS_3 with two-sided Z-score table
4.5.2	$ARMA(1,1)$ distance metric	1	Test statistic $TS_{4,\mathcal{T}^a}$ with chi-square distribution table
4.5.3.1	$INGARCH(1,1)$ residual	1	Test statistic TS_2 with two-sided Z-score table
4.5.3.2	$INGARCH(1,1)$ prediction	1	Test statistic TS_3 with two-sided Z-score table
4.5.4	$INGARCH(1,1)$ distance metric	1	Test statistic $TS_{5,\mathcal{T}^a}$ with chi-square distribution table
4.5.5	$INGARCH(1,1)$ intervention effect	1	**interv_test()** function
F.0.1.1	$ARMA(1,1)$ residual	1.a	Test statistic $TS_{2,\mathcal{T}^a,i}$ with two-sided score table with Bonferroni correction
F.0.1.1	$ARMA(1,1)$ residual	1.b	Test statistic $TS_{2,\mathcal{T}^a,i}$ with two-sided score table without Bonferroni correction
F.0.1.2	$ARMA(1,1)$ prediction	1.a	Test statistic $TS_{3,\mathcal{T}^a,i}$ with two-sided score table with Bonferroni correction
F.0.1.2	$ARMA(1,1)$ prediction	1.b	Test statistic $TS_{3,\mathcal{T}^a,i}$ with two-sided score table without Bonferroni correction
F.0.2	$ARMA(1,1)$ distance metric	1.a, 1.b	Test statistic $TS_{4,\mathcal{T}^a,i,j}$ with chi-square distribution table with Bonferroni correction
F.0.2	$ARMA(1,1)$ distance metric	1.c	Test statistic $TS_{4,\mathcal{T}^a,i,j}$ with chi-square distribution table without Bonferroni correction
F.0.3.1	$INGARCH(1,1)$ residual	1.a	Test statistic $TS_{2,\mathcal{T}^a,i}$ with two-sided Z-score table with Bonferroni correction
F.0.3.1	$INGARCH(1,1)$ residual	1.b	Test statistic $TS_{2,\mathcal{T}^a,i}$ with two-sided Z-score table without Bonferroni correction
F.0.3.2	$INGARCH(1,1)$ prediction	1.a	Test statistic $TS_{3,\mathcal{T}^a,i}$ with two-sided Z-score table with Bonferroni correction
F.0.3.2	$INGARCH(1,1)$ prediction	1.b	Test statistic $TS_{3,\mathcal{T}^a,i}$ with two-sided Z-score table without Bonferroni correction
F.0.4	Difference between $INGARCH(1,1)$ processes	1.a, 1.b	Test statistic $TS_{5,\mathcal{T}^a,i,j}$ with chi-square distribution table with Bonferroni correction
F.0.4	Difference between $INGARCH(1,1)$ processes	1.c	Test statistic $TS_{5,\mathcal{T}^a,i,j}$ with chi-square distribution table without Bonferroni correction
F.0.5	$INGARCH(1,1)$ intervention effect	1.a	**interv_test()** function with Bonferroni correction
F.0.5	$INGARCH(1,1)$ intervention effect	1.b	**interv_test()** function without Bonferroni correction

$TS_2 = \frac{\epsilon_{t_a}}{\hat{\sigma}}$.

$TS_3 = \frac{\epsilon_{t_a}}{\hat{\sigma}^*}$.

$TS_{4,\mathcal{T}^a} = T^a(\hat{\boldsymbol{\pi}}_x - \hat{\boldsymbol{\pi}}_y)^T \hat{V}^{-1}(\hat{\boldsymbol{\pi}}_x - \hat{\boldsymbol{\pi}}_y)$.

$TS_{5,\mathcal{T}^a} = \left\| \sqrt{\frac{T^a G_{\mathcal{T}^a}(\hat{\boldsymbol{\omega}}_a)}{2}} (\hat{\boldsymbol{\omega}}_1 - \hat{\boldsymbol{\omega}}_2) \right\|_2^2$.

$TS_{2,\mathcal{T}^a,i} = \frac{\epsilon_{t_{ia}}}{\hat{\sigma}}$.

$TS_{3,\mathcal{T}^a,i} = \frac{\epsilon_{t_{ia}}}{\hat{\sigma}^*}$.

$TS_{4,\mathcal{T}^a,i,j} = T^a(\hat{\boldsymbol{\pi}}_{(S_t)_{t\in\mathcal{T}^i}} - \hat{\boldsymbol{\pi}}_{(S_{t'})_{t'\in\mathcal{T}^j}})^T \hat{V}^{-1}(\hat{\boldsymbol{\pi}}_{(S_t)_{t\in\mathcal{T}^i}} - \hat{\boldsymbol{\pi}}_{(S_{t'})_{t'\in\mathcal{T}^j}})$.

$TS_{5,\mathcal{T}^a,i,j} = \left\| \sqrt{\frac{T^a G_{\mathcal{T}^a}(\hat{\boldsymbol{\omega}}_a)}{2}} (\hat{\boldsymbol{\omega}}_i - \hat{\boldsymbol{\omega}}_j) \right\|_2^2$.

Table E.3: A summary of all the time-dependent perspectives: A short description of the perspective with the tests per alternative hypothesis.

E.3. Figures of the β_i^j investigation per product team

Author's note: This figure is confidential.

Figure E.1: Two dimensional kernel density estimations of the non-zero values of $\frac{\beta_i^j}{\sigma}$ and $\frac{\beta_i^{j+1}}{\sigma}$ for every product team.

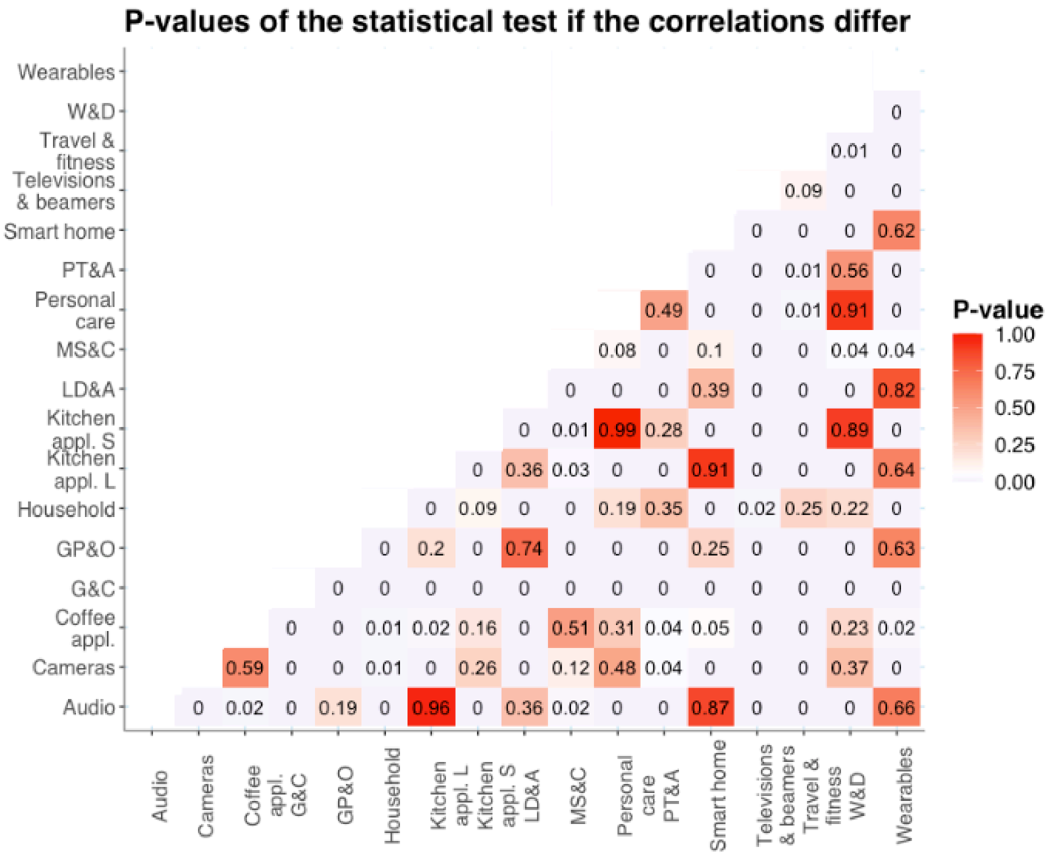

Figure E.2: P-values as an heat map of the test if the correlations differ significantly. Some on the product team names are abbreviated.

F

Time-dependent perspectives for products with multiple OOS periods

Perspective F.0.1: The value $s_{t_{ia}}$ is an outlier

- Perspective E.0.1.1: Difference between $\bar{s}_{t_{ia}}$ and $s_{t_{ia}}$

For $i = 1, ..., k+1$ let $(s_t)_{t \in (\mathcal{T}^i, \mathcal{T}^{i+1})} \sim ARMA(1,1)$ with fitted values $\bar{s}_t$, residuals $\epsilon_t = \bar{s}_t - s_t$ and $\sigma^i = sd(\epsilon_{t \in (\mathcal{T}^i, \mathcal{T}^{i+1})})$. Let c be a positive constant.

- $H_0 \,:\, \forall\, i \quad \frac{\epsilon_{t_{ia}}}{\sigma^i} \sim N(0,1),$
- $H_{1.a} \,:\, \exists\, i \quad \left|\frac{\epsilon_{t_{ia}}}{\sigma^i}\right| > c,$
- $H_{1.b} \,:\, \forall\, i \quad \left|\frac{\epsilon_{t_{ia}}}{\sigma^i}\right| > c.$

- Perspective E.0.1.2: Difference between $\hat{s}_{it_a}$ and s_{it_a}

For $i = 1, ..., k+1$ let $(s_t)_{t \in \mathcal{T}^i} \sim ARMA(1,1)$ with fitted values $\bar{s}_t$, residuals $\epsilon_t = \bar{s}_t - s_t$ and $\epsilon^*_t = \hat{s}_t - s_t$, and $\sigma^{*i} = sd(\epsilon_{t \in (\mathcal{T}^i, \mathcal{T}^{i+1})})$. Let c be a positive constant.

- $H_0 \,:\, \forall\, i \quad \frac{\epsilon^*_{t_{ia}}}{\sigma^{*i}} \sim N(0,1),$
- $H_{1.a} \,:\, \exists\, i \quad \left|\frac{\epsilon^*_{t_{ia}}}{\sigma^{*i}}\right| > c,$
- $H_{1.c} \,:\, \forall\, i \quad \left|\frac{\epsilon^*_{t_{ia}}}{\sigma^{*i}}\right| > c.$

Perspective F.0.2: Difference in coefficients between multiple $ARMA(1,1)$ processes

For $i = 1, ..., k+1$ let $(s_t)_{t \in \mathcal{T}^i} \sim ARMA(1,1)$.

- $H_0 \,:\, \forall\, i \in K, \quad d((s_t)_{t \in \mathcal{T}^i}, (s_t)_{t \in \mathcal{T}^{i+1}}) = 0,$
- $H_{1.a} \,:\, \exists\, i, j \in K, \quad d((s_t)_{t \in \mathcal{T}^i}, (s_t)_{t \in \mathcal{T}^j}) \neq 0,$
- $H_{1.b} \,:\, \forall\, i \in K, \quad d((s_t)_{t \in \mathcal{T}^i}, (s_t)_{t \in \mathcal{T}^{i+1}}) \neq 0,$

- $H_{1.c}$: $\forall\ i,j \in\ K, \quad d((s_t)_{t\in\mathcal{T}^i},(s_t)_{t\in\mathcal{T}^j}) \neq 0.$

with $d(\cdot,\cdot)$ the distance metric from Definition 4.5.4.

Perspective F.0.3: The value $s_{t_{ia}}$ is an outlier

- Perspective E.0.3.1: Difference between $\bar{s}_{t_{ia}}$ and $s_{t_{ia}}$

For $i = 1,...,k+1$ let $(s_t)_{t\in(\mathcal{T}^i,\mathcal{T}^{i+1})} \sim INGARCH(1,1)$ with fitted values $\bar{s}_t$, residuals $\epsilon_t = \bar{s}_t - s_t$ and $\sigma^i = sd(\epsilon_{t\in(\mathcal{T}^i,\mathcal{T}^{i+1})})$. Let c be a positive constant.

- H_0 : $\forall\ i \quad \frac{\epsilon_{t_{ia}}}{\sigma^i} \sim N(0,1),$
- $H_{1.a}$: $\exists\ i \quad \left|\frac{\epsilon_{t_{ia}}}{\sigma^i}\right| > c,$
- $H_{1.b}$: $\forall\ i \quad \left|\frac{\epsilon_{t_{ia}}}{\sigma^i}\right| > c.$

- Perspective E.0.3.2: Difference between $\hat{s}_{it_a}$ and s_{it_a}

For $i = 1,...,k+1$ let $(s_t)_{t\in\mathcal{T}^i} \sim INGARCH(1,1)$ with fitted values $\bar{s}_t$, residuals $\epsilon_t = \bar{s}_t - s_t$ and $\epsilon_t^* = \hat{s}_t - s_t$, and $\sigma^{*i} = sd(\epsilon_{t\in(\mathcal{T}^i,\mathcal{T}^{i+1})})$. Let c be a positive constant.

- H_0 : $\forall\ i \quad \frac{\epsilon^*_{t_{ia}}}{\sigma^{*i}} \sim N(0,1),$
- $H_{1.a}$: $\exists\ i \quad \left|\frac{\epsilon^*_{t_{ia}}}{\sigma^{*i}}\right| > c,$
- $H_{1.c}$: $\forall\ i \quad \left|\frac{\epsilon^*_{t_{ia}}}{\sigma^{*i}}\right| > c.$

Perspective F.0.4: Difference in coefficients between multiple $INGARCH(1,1)$ processes

For $i = 1,...,k+1$ let $(s_t)_{t\in\mathcal{T}^i} \sim INGARCH(1,1)$ with ω^i the parameter vector.

- H_0 : $\forall\ i \quad \omega^i = \omega^{i+1},$
- $H_{1.a}$: $\exists\ i,j \quad \omega^i \neq \omega^j,$
- $H_{1.b}$: $\forall\ i \quad \omega^i \neq \omega^{i+1},$
- $H_{1.c}$: $\forall\ i,j \quad \omega^i \neq \omega^j.$

Perspective F.0.5: Intervention effect

For $i = 1,...,k+1$ let $(s_t)_{t\in\{\mathcal{T}^i,\mathcal{T}^{i+1}\}} \sim INGARCH(1,1)$ with an intervention at time t_{ia}.

- H_0 : $\forall\ i \quad \eta_{t_{ia}} = 0,$
- $H_{1.a}$: $\exists\ i \quad \eta_{t_{ia}} \neq 0,$
- $H_{1.b}$: $\forall\ i \quad \eta_{t_{ia}} \neq 0.$