UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering,
Mathematics & Computer Science

Master Thesis Applied Mathematics
Stochastic Operations Research Group

Modeling an optimal promotion strategy for FMCG adjusted for cannibalization effects

Fabienne S. Mouris

April 2023

DATA

Supervisors:
dr. J. C. W. van Ommeren
drs. J. van Dijk (D-Data)
dr. ir. V. de Graaff (D-Data)

Graduation Committee:
prof. dr. R. J. Boucherie
dr. J. C. W. van Ommeren
dr. M. Walter
drs. J. van Dijk (D-Data)
dr. ir. V. de Graaff (D-Data)

Preface

This thesis completes my master's program in Applied Mathematics at the University of Twente which I started in September 2020 after retrieving my Bachelor's diploma in Econometrics. For the past 8 months, starting from September 2022, I have been dedicatedly working on this project, encountering various challenges. The successful completion of this thesis would not have been possible without the generous assistance of many people.

The warm welcome at D-Data to guide me during the final part of my studies marks the start of this thesis. Their guidance and supervision supported me during the challenges of this period, while also making it fun, interesting, and highly informative. I would like to thank all my colleagues who are always open to discussing the method or playing a game of table soccer to relieve some stress. In particular, I would like to express my sincere graduate to Victor de Graaff and Jeroen van Dijk for offering me to the opportunity to write my master's thesis at D-data and their valuable supervision. I truly appreciate Jeroen's daily supervision, who has spent a lot of time and effort providing insightful feedback on the method, coding, and report.

Moreover, I would like to extend my gratitude to Jan-Kees van Ommeren my supervisor from the University of Twente for his constructive feedback, flexibility, and visits to 's Hertogenbosch. Also, I would like to thank my graduation committee consisting of Richard Boucherie, Jan-Kees van Ommeren, and Matthias Walter for the assessment of my thesis.

Finally, I want to acknowledge the support, motivation, and occasional distraction from my friends and family throughout this period. I want to express my special thanks to Olivier for his mental support and for proofreading my thesis.

I look back with great pleasure on my time as a graduate intern and look forward to joining D-Data in August.

Fabienne Sarita Mouris

Abstract

This research has developed a model for an optimal promotion strategy for FMCG adjusted for cannibalization effects. A regression model is developed that estimates the cannibalization effects and returns a sales volume forecast based on promotion variables and their lagged value. Subsequently, the regression model is used to define the promotion calendar optimization problem that returns an optimal promotion calendar while maximizing the total sales volume of all products. The logistic regression is evaluated by calculating the MAPE and R^2 indicating a great deviation in absolute volume while the R^2 indicates an accurate fit. Therefore, the categorical R^2 is developed to accurately measure the performance of regression models that deal with clustered data. Results show that the estimated coefficients are not fitting the data. Nevertheless, the sensitivity analysis of the regression coefficients reveals that they do not significantly impact the optimization problem. This indicates that the regression model is suitable for analyzing the promotion calendar optimization problem within the constraints of the business if the cannibalization effects are depicted correctly and do not change drastically. Results show that there exists a promotion plan with an increased sales volume for the FMCG chain and that more promotional weeks do not necessarily result in the highest turnover.

Keywords: sales forecast, logistic regression, cannibalization effects, promotion strategy, optimization, coefficient of determination

Contents

Chapter 1

Introduction

In the past years, the market of fast moving consumer goods (FMCG) has been growing rapidly. The FMCG market is competitive and faces several challenges. This market is a dynamically changing environment with great product diversity but within each segment great substitutability between products. As a result, there is intense competition between companies operating in the FMCG market which shows characteristics of perfect competition. The FMCG market is a classic example of low-margin and high-volume sales.

In the FMCG market, sales are driven by the level of the price. Due to varying price elasticity, companies often have to revise the price to keep their profit maximized. Besides this, price adaptions can be driven by the limited life cycle of products or seasonal effects like the weather, holidays, and other festivities. Especially non-perishable products such as beer and alcoholic drinks are sensitive to these effects and a good pricing policy is found to be a powerful marketing strategy [33]. Consequently, FMCG companies offer sales promotions, featuring discounts of up to 50% or more.

However, if one company reduces the price, competing companies selling similar products will do the same because otherwise, they will lack sales and market share, even if it is a temporary discount only. Although the companies would be better off without any price reduction, they still offer it for a low price which makes it a typical example of the prisoner's dilemma[1]. The prisoner's dilemma illustrates the conflict between the individual and collective rationality [23].

FMCG companies compete in the pricing battle with substitute goods when the effects of sales promotion are profitable. Despite discounting the price, the revenue should increase. Hence, the volume turnover should inflate. Assuming that the demand within the market cannot inflate as much as needed to reach the volume

[1]Prisoner's dilemma first raised by M. Flood and M. Dresher is a classic game where two individuals acting in their interest with a sub-optimal outcome. If the individuals would cooperate, their profit would be higher compared to the situation when they did not. However, there is a risk the other individual deviates from cooperating. Then, the individual deviating ends up with all profits, and the one willing to cooperate with less profit, even in the case where both parties defect.

turnover needed, the sales will replace from one company to another. The effect of consumers switching products as a result of a change in the price is called cannibalization.

Although pricing promotions are not the most profitable for companies, they must develop a strategy that maximizes the sales volume and market share. That is also the case for FMCG companies in the beer market. One of the largest beer breweries in the Netherlands, and an FMCG company, is Heineken and it carefully ponders its marketing strategy and price promotions at the retailer because they have to consider its own products next to cannibalizing products of contenders. Due to great product assortment and its diversity, decision-making on pricing strategy is becoming more complicated while the price is a key element of the marketing mix because of fierce competition and dynamically changing environment [33]. Accurate sales forecasts and optimal promotion planning are crucial for FMCG companies.

Hence, this research paper will focus on setting up a pricing strategy for an FMCG company that takes into account the promotion effects of its own cannibalizing products and competitors. The model will return a calendar that advises the company when to sell which products for a discount price while a forecasting formula is included to maximize the sales. Sales forecasting is widely researched but its accuracy is different across markets. Therefore, a case study on retail data from the beer market is discussed. The sales predictions are used as input for the model that optimizes the promotion calendar. Research on optimization models that incorporate cannibalizing effects is limited while an accurate model is highly valued in the FMCG market.

In Chapter 2, a theoretical background is provided about the topics, cannibalization, pricing strategies, and sales increase due to promotions. In the second part of this chapter, background information about the company D-Data is described, continued with the research goal and its motivation. Since this research combines two models, Chapter 3 discusses the literature on econometric models for sales forecasting and presents the mathematical model for forecasting. Chapter 4 continues with the second model on optimizing the promotion calendar by maximizing the sales volume. The results of the case study on retail data of the beer market are presented in Chapter 5. Finally, the conclusion and discussion can be found in Chapter 6.

Chapter 2

Problem description and approach

This chapter provides background information regarding the FMCG market, cannibalization, and the role of pricing. The second part of this chapter elaborates on the research performed, the research goal, and the company where this research is pursued.

2.1 Theoretical background information

Supermarkets are retailers that sell a great diversity of FMCG produced by FMCG manufacturers. These products owe their name to their short shelf life. Other characteristics are that products are purchased frequently for a low price in large quantities, and are consumed rapidly. A product is related to one brand and one manufacturer. Moreover, multiple products can be produced under the same brand name and a manufacturer may have introduced multiple brands. Manufacturers have introduced multiple product lines and many companies are competing in the market. A product line of a manufacturer refers to a group of products of the same brand with similar characteristics. Each manufacturer invests in marketing and new product lines to make sure their brand is well known and has a successful position on the market.

2.1.1 Cannibalization effects

Cannibalization occurs in markets with a great product assortment. Cannibalization is the phenomenon of a product eating sales or market share of a substitute. The conditions for cannibalization are, the products should be similar, quality can not differ too much and the price gap is large enough to trigger shifts in demand [29]. Consumers' decision to purchase a product might depend on marketing and price discounts. They switch between products such that the sales from one product decrease while sales of another product increase. Products with similar characteristics cannibalizing each other are collected in a product group.

Cannibalization is a challenging aspect within demand forecasting, inventory purchasing, and pricing strategy determination. Moreover, it is often neglected since it is complex to measure cannibalization effects and even harder to include it in optimization algorithms [40]. Ivanov [21] focuses on defining types of cannibalism and setting up a conceptual framework to analyze cannibalization. There are two levels on which cannibalization occurs, company- and industry-level. Product innovation might cause cannibalization on company level. Conversely, cannibalization on industry level is based on competition for market share where the larger firm wants to knock the smaller one off the market to cut down the competition in the industry. The most common type to investigate is product cannibalization which occurs on company level. When a new product is introduced by a company, this could cause an increase in the sales and profit of the new product at cost of the older product(s) already on the market. However, this does not only occur within companies but also across. Hence, the research about product cannibalization on company- and industry-level is often combined.

It is well known that a great variety of products increase the demand and market share of a company. However, this causes self-competition and hence product cannibalization. Namely, the newly introduced product competes with the already existing products. Since consumers are heterogeneous in their choice of taste and quality, this is an effect manufacturers should consider [31]. If products have more characteristics in common, the possibility becomes higher that a new product eats sales from a competing product or cannibalizes an existing product [29]. The sequential introduction of products is an option to deal with the cannibalization of similar products. Mainly for technological items it is profitable to sequentially introduce the products. However, the manufacturer has to consider whether the postponement of the revenues is advantageous in comparison to incurring the cannibalization effects. The cost of later introduction can be higher than the yields of reducing the cannibalization effects. Therefore, the expected expenses and yields in both scenarios are investigated by the manufacturer to decide on the sequential or simultaneous introduction of products [30].

Faria and Novak [12] have developed a dynamic model for corporate cannibalism which can be used to analyze an oligopolistic industry, that is an industry consisting of a small number of firms that have significant domination of the industry. Since they have defined cannibalism as extremely aggressive behavior towards competitors, the purpose of larger firms is to drive smaller competitors out of the market and keep new competition from entering the market. As a result, the larger firm's market share and profit increase. However, in a cannibalizing oligopolistic industry, the number of firms fluctuates according to a biological model. First, larger firms have a more efficient production process, hence reduced costs leading to fewer rivals on the market. This gives the large firms the opportunity to increase their degree of monopoly and increase the price which attracts new entrants to the market. The cycle will repeat itself which explains the fluctuation in the number of firms in an oligopolistic industry.

In a large assortment of products, it is better for a company that the consumer chooses between one of their products rather than one of their competitors. Especially, since companies have multiple brands in their portfolio and thus products within the same category [26]. In the beer market, there are manufacturers owning multiple brands, for example, the famous Dutch brewery Heineken. They started off with the production of Heineken but also introduced Amstel and the non-alcoholic beer brand Heineken 0.0. Besides, they renowned Tiger, Desperado's, and Birra Moretti while owning several cider and hard seltzer brands [17]. Each brand could contain products from different product categories. For instance, special beers IPA, Bock, Blond, or Triple of Brand and Affligem are also brands of the manufacturer Heineken. This is in line with the recent growth of markets to introduce new items in a product line under the same brand since creating a brand and giving it a new identity does take time and is financially costly. Moreover, it is not rare to find a large number of alternatives within a single product segment. This enlargement results in cannibalization which ideally eats sales from a product of another manufacturer, but the cannibalization effects on an assortment of their own company have to be considered [26].

Mason and Milne [26] developed a method to identify cannibalization between items within a brand or between brands of a manufacturer. The method analyzes pairwise cannibalization of items within brands and total cannibalization at brand and manufacturer level. It is an ecological framework where each brand has a niche describing the customers the brand is competing for. The dimensions of the niche include customer characteristics and vary per product segment. A customer belongs to a brand's niche and is a core customer if it buys the brand it belongs to. In case a customer purchases a brand it does not belong to, the potential cannibalization can be analyzed. Next to negative effects, cannibalization could hinder competitors to increase their sales. In other words, cannibalization is a result of competition in a fragmented market.

2.1.2 Product pricing

Cannibalization is influenced by introducing line extension and marketing policies within the company, but also external factors might cause cannibalization, e.g. changes in the price or the introduction of a new product by competitors. The more products are comparable, the stronger the cannibalization effects are. Given the importance of price in the marketing mix, understanding the impact of price on cannibalization is valuable to companies. Meredith and Maki [29] researched the role of price between two brands where the lower-priced brand cannibalizes the premium brand. The introduction of the lower-priced brand resulted in a decrease in the total sales volume of the higher-priced brand. In this paper, a log regression is performed to predict the market share of brand 'B' while measuring the effect of the price gap of premium brand 'A' and brand 'B' and including a lag. The research concludes that low-priced brand 'B' eats the market share of premium brand 'A', and so could other cheap substitutes. The price of a product plays a significant role in the cannibalization effect.

Cannibalization occurs not only between lower-priced products and premium products but also a temporary change in the price, a price promotion, of a product might cause an increase in sales at cost of similar products. In fact, it shows that cannibalization helps to prevent spoilage and out-of-stock moments of companies, leading to better capacity, utilization, and higher revenue [28].

The research by R. C. Blattberg and Wisniewski [4] claims that promotional price competition results in asymmetric sales effects regarding products having similar attributes except for price and quality. Premium products with a higher price and quality eat sales from competing brands with similar or lower price products with at most the same quality. Contrarily, it is less likely that lower price products draw sales of premium products with higher quality. However, when promotional discounts are extreme, lower-priced brands can affect premium brands when consumers find the compensation in form of a discount for the lower quality sufficient. Still, competitive effects may be asymmetric. One brand can be a strong competitor of another brand while the other brand may have no or little effect on the first one.

2.1.3 Sales promotion models

In the second half of the 20^{th} century, sales promotions became more common for consumer goods. Still, the modern grocery retail environment is severely influenced by price promotions which is a major component of the competitive FMCG market. A price reduction or sales promotion is established for increasing purchase frequency and impulsive purchases. Furthermore, it attracts disloyal consumers. Nonetheless, price promotions have a direct but temporary influence on sales volumes [6],[18]. The effects are measured by descriptive models. Furthermore, prescriptive models use the outcome of descriptive models to generate advice on how to use or improve sales promotions. This can be investigating what price would be beneficial or when to introduce a price reduction. Hereby, an optimization model using mathematical programming is used to maximize the profit, the objective function [5].

Research on descriptive models mainly focuses on the profitability of price promotions. Greenleaf [15] investigated this in 1995 from retailers' perspective to help them formulate a more profitable promotion strategy for a single period. It is concluded that irregular promotions are profitable for the retailer. Moreover, the model is helpful for manufacturers to convince retailers that pricing promotions are beneficial. More recent research is done by Hosseini et al. [18]. They found that price reductions have a significant positive effect on impulsive purchase behavior which concludes that promotions are profitable. The focus of McColl et al. [27] and Van Heerde et al. [38] is whether the sales bump is profitable for the manufacturer and explores the cross-brand price effects. Cannibalization, and thus brand switching, is not the only result of promotion pricing but also stockpiling and increased consumption are consequences. From the retailer's perspective, brand switching is not so interesting but store switching is. That is why sales bump is split into cross-brand, cross-period, and category-expansion effects. The effects are estimated by using regression models depending on promotion binary variables of the product at several points in time which include lagged and shifted variables to measure the effect of the promotion. In stable environments, on average brands' net loss are

33% of their sales to price-promoted brands [38]. The amount of price discount does influence the effect measured. The cross-brand and cross-period effects have little impact while the category-expansion effect increases for higher price discounts. Also, the type of store has an impact on the cannibalization effect. Large stores deal with stronger cannibalization effects than medium and small stores.

2.1.4 Promotion planning in the Netherlands

Given the aspects involved in the role of prices and planning promotion strategies, this section will provide insights into the process of assigning promotion slots to products in retail stores. Manufacturers have to purchase the promotion slot from retailers, but ultimately it is up to the retailer to decide upon the promotion week. Therefore, the manufacturer can only choose to either accept or decline the promotion slot offered by the retailer. However, the FMCG market is competitive, and since manufacturers aim to achieve the highest market share, they will not give away promotion slots to competitors. It is important to note that the manufacturer with the largest market share has priority in negotiating with the retailer regarding promotion slots, followed by the second and third largest. Only the three largest manufacturers in terms of market share have the opportunity to bargain with the retailer since it is a time-consuming process. The retailer aims to expand the market and achieve growth in total sales while the manufacturer's priority is to increase its turnover. Thus, when proposing a promotion plan to the retailer, manufacturers also take into consideration the retailer's objectives alongside their own. Typically, a promotion plan is proposed for an entire calendar year where each promotion slot represents one week.

The manufacturer does not only propose a promotion plan but also considers for which price the product is promoted. There are two types of promotions: deep and undeep. A deep promotion involves a discount of at least 40%, while an undeep promotion offers a discount ranging between 10% and 40%. The degree of discount affects both the cannibalization effects and sales volume, hence it is crucial to choose the discount level strategically.

2.2 Problem definition

With the background information in mind, this section describes the problem, and the research question is defined. The approach and layout of the thesis will be elaborated on.

2.2.1 The company: D-Data

D-Data (https://www.d-data.nl/) is a small and young consultancy company located in 's Hertogenbosch. They focus on all components of Data Science and Engineering. The way of working at D-Data consists of 6 steps:

1. problem definition;
2. data gathering, cleaning and visualizing;
3. model development;
4. model review by colleagues;
5. model validation and industrialization, and;
6. monitoring, governance and documentation.

D-Data works together with clients such as Unilever, Heineken, and Vrumona. D-Data fulfills projects in many sectors among them are retail, finance, and non-profit organizations. Besides working for different clients, D-Data also has its own internal projects. One of the projects that five employees collaborated on is SalesBrain (https://demo.salesbrain.ai/).

2.2.2 Company project: Salesbrain

Salesbrain is a web application built in Power BI, which is developed for the sales data of a retail store with several store locations. It provides insights into sales data divided into 5 different topics:

- overview of historical sales transactions
- local product performance
- temperature effects
- price elasticity effects
- cannibalization effects

The user has the possibility to retrieve information per store, brand, or product selection. The local product performance compares the sales in a store with similar stores in the same regions. The cannibalization effects are determined by comparing the weekly sales of a product during a promotion against the sales of another product in the same segment that is not promoted. Although measuring the cannibalization effects requires research, this is a first and simple idea to assess absolute cannibalization effects.

2.2.3 Research goal and motivation

The simple method to assess cannibalization effects raises the aspiration to further research cannibalization effects in the FMCG market for frequently promoted products with a lot of substitutes. First, D-Data analyzed various techniques to identify cannibalizing products within the same category due to promotional offers. D-Data previously researched cannibalization effects and found that Vector Auto Regression (VAR) model is useful for identifying cannibalizing products or product groups. This model uses multiple variables and a predetermined number of lags of the dependent variable to predict the dependent variable. The reason this model was found best is that it does not matter if variables influence each other. So, the exogenous variables may be correlated. However, the forecast accuracy is on average 34% which is not sufficient for FMCG companies. D-Data is looking for a method that gives more accurate cannibalization effects between products. In other words, a method that estimates the cannibalization effect which is closer to the actual effect than the currently predicted effect.

Therefore, this research develops a model that identifies cannibalization effects as a result of price discounts on similar products in the same product category. To obtain insights into the forecast accuracy of the model on clustered data, a categorical coefficient of determination is designed. The main goal is to suggest an optimal promotional strategy across time for cannibalizing products in the same category of multiple FMCG companies by using the estimated cannibalization effects. In this case, optimal refers to maximizing the sales volume.

To achieve this goal, the research is split into two parts. In the first part, Chapter 3, a sales forecasting model is developed that predicts the sales volume considering the cannibalization effects of cannibalizing products. The estimated sales volume is presented by a multiplicative prediction formula taking into account seasonal and cannibalization effects if a cannibalizing product is promoted. This formula is linearized by log transformation such that ordinary least squares can be applied for parameter estimation. Moreover, the method for the categorical coefficient of determination is explained. In the second part, Chapter 4, an optimization model is developed that decides for each product which time periods it should be offered for a discount price. The objective function is to maximize the sales volume of the FMCG chain for which the prediction formula from the first part is used. The FMCG chain includes the retailer and the manufacturer such that the manufacturer and the retailer have their sales maximized.

Although the competitive FMCG market shows characteristics of the prisoner's dilemma, the problem is not analyzed with game theory models on non-cooperative games. For each manufacturer, the best response functions should be set up taking into consideration all products the manufacturer produces. Besides, the competitor's strategies are assumed to be given which is in practice uncertain since there are many strategies a manufacturer can choose. Also, the cannibalization effects should be estimated and incorporated into the utility function of the manufacturers. Forecasting via linear regression estimates cannibalization effects simultaneously. The optimization is not only interesting for the manufacturers but certainly for the retailer. Therefore, this research focuses on sales prediction via linear regression models and optimization via mathematical programs.

Chapter 3

Sales forecasting by econometric regression models

This chapter gives an overview of former literature research and presents the mathematical method for the econometric regression model. Then, the regression model is validated by statistical testing, and forecast accuracy measures are defined.

3.1 Literature research

Sales forecasting is important for FMCG manufacturers. They want to conquer the market and therefore, the company's reputation, financial situation, and position in the market should be better than those of competitors. This can be improved by being more sustainable. Besides overproduction being unsustainable, it is very costly. So inaccurate demand planning is the origin of uncertainty at FMCG companies which causes high costs due to overproduction. Therefore, better forecast accuracy than competitors can improve efficiency within the supply chain and gain an advantage on market [3].

Forecasting methods use historical data to identify trends and current knowledge to predict future market trends. Quantitative data, which are represented as drivers of the outcome, are transformed in order to help businesses in their decision making process. Especially econometric techniques come in handy to use factors affecting retail sales for its prediction. These factors are seasonal effects such as holidays, festivals, and other events, as well as external elements, e.g. weather and pandemics. Moreover, they influence the sales volumes but also price discounts and promotions for own and competitor products [16], [32].

In former years, regression and time series models have been used for forecasting. To analyze these models, techniques such as simple linear regression and exponential smoothing toward more complex VAR models were applied. These models accurately find trends, seasonal effects, and auto-correlation under the assumption that price and promotional effects are constant [1]. Structural change is not addressed in retail sales forecasting most of the time before the 2000's [20]. Conventional models predict sales which assume that the future follows the exact pattern as the past. Data about economic conditions, changes in consumer behavior, and new entrants in the market are poorly available. Hence, these structural changes are omitted thus far. Then, Huang et al. [20] was the first to consider structural changes and has proven to retrieve more accurate sales predictions than conventional models.

The structural change and the presence of promotions within the FMCG sector made sales forecasting more complex, yet more important. A complex model is ACNielsen's SCAN*PRO model which shows the effects of momentary price reductions based on price discount elasticity, seasonal trends, and cross-brand effects. Van Heerde et al. [37] discuss extensions of the SCAN*PRO model with a focus on a relatively simple initial model without including all complexities to facilitate acceptance by managers, but containing enough data information to improve managers' decisions. The model is adjusted for dynamic effects and raised in complexity by including leads and lags for price promotion variables to accommodate pre- and post-promotion dips in sales data. Decomposing the sales data is a convenient approach to recognize the sales bumps on the brand level due to promotion and provides insights into the brand competition by applying econometric models. From qualitative data the exogenous variables, or features, are determined. These features may include summary statistics, but also cross effects, e.g. products and powers, such that they provide sufficient information to forecast the dependent variable by a linear relation or other machine learning techniques [32]. According to Heij et al. [16] a linear regression is a method of estimating the conditional expected value of the dependent variable given the values of a set of predictor variables. Ideally, the parameters of the explanatory are a robust estimate for good forecast accuracy and therefore elimination of outliers may be necessary.

Promotional forecasting is relevant for retailers and even more for manufacturers of FMCG. Rather than looking into the absolute number of units, the percentage increase is of interest. Common within promotion forecasting are synergy effects between exogenous variables, amongst which are cannibalization effects. For example, imagine three cannibalizing products of which two are promoted in the same week. Then the negative effect on the sales volume of the not promoted product in that week is stronger than the sum of the individual effects of the promoted products. Hence, multiplicative regression models are often used for promotional forecasting. Parameter estimation can still be done according to linear regression techniques after logarithmic transformation of the data [32].

Further literature suggests using Multi Linear Regression (MLR) when the correlation is assumed linear between the predictors and the dependent variable. Moreover, the predictors cannot be correlated to prevent biased parameters and non-robust models. Price is the most important predictor for sales and is widely used by researchers, followed by cannibalization effects. Next to historical data, external factors weather, and holidays are included as predictors. When the mean absolute percentage error (MAPE) is low and the R^2 - the variation of variability - is high, the MLR is a suitable model to predict demand [13]. Furthermore, the error terms must be normally distributed with mean zero, constant variance, and no auto-correlation to guarantee non-biased parameter estimations.

Predictors of logistic regressions and MLRs are defined by available data, such that cannibalization effects can present an independent variable in the regression. Cannibalization can be measured company-wide, having only one predictor. On the other hand, it can be measured on the product level where each product has its own cannibalization effect on the dependent product variable. Therefore, an Autoregressive Distributed Lag (ADL) model can be used for predicting sales per product [19], [20]. This model considers lagged values of the dependent as well as the independent variables. The number of lags for each of the variables does not need to be the same. Also, a number of lags of the error terms could be added to the model. The Schwarz Information Criteria (SIC) and Breusch Godfried test can be used to test the significance of and the auto-correlation between the number of lags [16]. Huang et al. [19] have set up an ADL regression where the explanatory variables are determined from their own company's and competitors' historical data. The goal is to forecast the sales of a product which depends on the product price and product promotional weeks, as well as the competitors' product price and promotional weeks. Additional predictors are holidays and seasonal effects where the seasonal effect is split into four-week periods. Each four-week period is presented by a dummy variable. In this research, at most two lags of price and promotional variables are included. In comparison to the benchmark models of this research, the model with two lags has a higher forecast accuracy. A side note is that the model is assumed to be stationary. Therefore, influencing factors, e.g. economical conditions or consumer behavior, are not taken into account. This is especially noticeable in long-term forecasting.

3.2 Methodology

To achieve the main goal defined in 2.2.3, the sales forecasting model has to be developed which accurately predicts the sales per product of own company and competitors using historical data. This section will describe how the forecasting formula is defined. First, preliminary variables and regressors and the model assumptions are defined, then the regression formula is presented.

3.2.1 Preliminaries and regressors

Forecasting results are retrieved on a time horizon $|T|$, where T is the set of time points. Besides, consider M the set of manufacturers and P_m the set of products belonging to manufacturer $m \in M$. P_m is a subset of P containing all cannibalizing products of all manufacturers. The goal is to forecast the sales of a product at a given point in time. That is why the actual sales values are stored in y_{it} for product $i \in P$, $t \in T$, and the predicted values are presented by $\hat{y}_{it}$. The literature in Section 3.1 suggests including regressor variables that drive the sales, which are the median sales of the forecasting product $\tilde{y}_i$ and the seasonal weekly effect π_t which holds for the products in the same segment. Since the FMCG market has a large assortment of substituting products created by a great variety of manufacturers frequently offering their products for a promotional price, the cannibalization effects due to promotional offers of all cannibalizing products are included in the regression by dummy variables. A dummy variable is a binary variable that equals 1 if the product is offered for a promotion price and 0 otherwise. Obviously, the promotional variable is product and time-dependent. Hence, a promotional variable is denoted as x_{jt} for all $i \in P$, $t \in T$. Since promotional offers have a post-sales dip, one lag of the promotional variable is added to the regression model such that the covariate effects are accounted for.

When determining the independent variables for the regression formula, the correlation between these has to be checked by several tests. The variables should be independent to prevent multicollinearity and inflated error terms which makes it difficult to statistically test the regression coefficients. The Pearson correlation coefficient [34] can be used to test the correlation between continuous variables and the Chi-squared test is appropriate to test the correlation between categorical variables and or the combination of categorical and continuous variables.

3.2.2 Regression formula for sales prediction

The relation between the regressors mentioned in the previous section is not necessarily linear due to the presence of synergy effects. Hence, a multiplicative regression formula is set up which can be used for sales predictions. The regression formula defining the sales volume is

$$\nu_{it} = \pi_t^{\gamma_{1i}} \tilde{y}_i^{\gamma_{2i}} \prod_{j \in P} \beta_{1j}^{x_{jt}} \beta_{2j}^{x_{jt-1}} \varepsilon_{it}$$

where γ_{1i}, γ_{2i} for $ii \in P$, and β_{1j}, β_{2j} for $j \in P$ are the estimated coefficients.

A multiplicative regression formula is transformed by the log function in order to linearize the relationship between the independent and dependent variables and estimate the coefficients using ordinary least squares (OLS). Since this research deals with sales data and the regressors all have variables greater than zero, the log function is defined continuously on the domain of the variables. The log-transformed result is

$$\begin{aligned} \log(\nu_{it}) = {} & \gamma_{1i}\log(\pi_t) + \gamma_{2i}\log(\tilde{y}_i) \\ & + \sum_{j\in P} \log(\beta_{1j})x_{jt} + \log(\beta_{2j})x_{jt-1} \\ & + \log(\varepsilon_{it}). \end{aligned} \tag{3.1a}$$

In formal notation, the predicted value $\log(\nu_{it})$, the estimators $\log(\beta_{qj})$ and error term $\log(\varepsilon_{it})$ are respectively denoted as $\hat{y}_{it}$, b_{qj} and e_{it} where $i, j \in P$ the products and q the index of the lag variable. Then, the simplified prediction formula is

$$\begin{aligned} \hat{y}_{it} = {} & \gamma_{1i}\log(\pi_t) + \gamma_{2i}\log(\tilde{y}_i) \\ & + \sum_{j\in P} b_{1j}x_{jt} + b_{2j}x_{jt-1}. \end{aligned} \tag{3.1b}$$

Notice that the difference between the actual value and the predicted value is the error term. Besides, we denote the log transform of the actual volume, $\log(\nu_{it})$, as y_{it}. Hence, Equation (3.1a) can be written as

$$y_{it} = \hat{y}_{it} + e_{it}.$$

The regression formula in matrix notation for product i is

$$\mathbf{y}_i = X_i\mathbf{b}_i + \mathbf{e}_i, \tag{3.2}$$

with

$$\begin{aligned} X_i = {} & \begin{bmatrix} \log(\pi_2) & \log(\bar{y}_i) & x_{12} & \dots & x_{|P|2} & x_{11} & \dots & x_{p1} \\ \log(\pi_3) & \log(\bar{y}_i) & x_{13} & \dots & x_{|P|3} & x_{12} & \dots & x_{p2} \\ \vdots & & & \dots & & & & \vdots \\ \log(\pi_{|T|}) & \log(\bar{y}_i) & x_{1,|T|} & \dots & x_{|P|,|T|} & x_{1,|T|-1} & \dots & x_{|P|,|T|-1} \end{bmatrix} \\ & \in \mathbb{R}^{|T|-1\times K}, \end{aligned}$$

and

$$\begin{aligned} \mathbf{b}_i^T = {} & \begin{bmatrix} \gamma_{1i} & \gamma_{2i} & \log(\beta_{11}) & \dots & \log(\beta_{1|P|}) & \log(\beta_{21}) & \dots & \log(\beta_{2|P|}) \end{bmatrix} \\ & \in \mathbb{R}^K, \end{aligned}$$

where

$$\mathbf{e}_i = \begin{bmatrix} e_{i2} & e_{i3} & \dots & e_{i,|T|-1} \end{bmatrix} \in \mathbb{R}^{|T|},$$

and

$$\mathbf{y}_i = \begin{bmatrix} y_{i2} & y_{i3} & \dots & y_{i,|T|-1} \end{bmatrix} \in \mathbb{R}^{|T|}.$$

The time of the vectors and matrix in Equation (3.2) starts from $t = 2$ because one lag of the promotional variables is added to the regression model.

To estimate the parameters of the regression formula, the OLS is applied. This technique minimizes the sum squared residual and retrieves parameter estimates by setting the first-order derivatives equal to zero. A full derivation is given in Appendix A. Chapter 5 present the data used for experimental results and parameter estimation which is described in Section 5.3.1.

3.2.3 Model validation

Next to the independency assumption for the regressors and the linear relation on the log-transformed regression formula, there are assumptions that should hold for the error term. The error term, or residual, is the difference between the predicted and actual value. The error terms should have mean zero, homoskedastic variance σ^2, and non-autocorrelation. In addition, they should fit a random normal distribution. This results in

$$e_i \sim N(0, \sigma^2 I).$$

Test for normal distribution

The Kolmogorov-Smirnov (K-S) test [24] is used to verify if the sample distribution is a good fit for the statistical distribution. The K-S test is used for continuous distributions and is based on the maximum distance in the vertical direction of the cumulative sample distribution and cumulative theoretical normal distribution function.

Definition 1 (K-S test)**.** Consider the set of errors $\mathbf{e} = \{e_1, e_2, \ldots, e_{|T|}\}$ and theoretical standard normal distribution function $F(x)$. The mean and standard deviation of the set $\mathbf{e}$ are 0 and s such that the normalized error set is $\frac{\mathbf{e}}{s}$. Define

$$\begin{aligned} H_0 &: \frac{\mathbf{e}}{s} \xrightarrow{d} N(0, I) \\ H_a &: \frac{\mathbf{e}}{s} \text{ not standard normally distributed.} \end{aligned}$$

The K-S statistic is defined as

$$D = \max_{x \in \mathbb{R}} F_{X,n}(x) - F(x),$$

where the cumulative sample distribution $F_{X,n}(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{x_i \leq x\}$, i.e. the fraction of sample observations with a value smaller than or equal to x.

If test statistic D is greater than the critical value, H_0 is rejected.

Test for homoskedasticity

The error vector is homoscedastic if its random samples have the same finite variance. The most popular test for homoskedasticity is Bartlett's test [2]. This test examines whether the variance of the dependent variable is the same for different groups.

Definition 2 (Bartlett's test)**.** Consider the N data points clustered into k groups such that the i^{th} group, consisting of n_i data points, has variance σ_i^2, for $i \in K = \{1, k\}$. Define

$$\begin{aligned} H_0 &: \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 \\ H_a &: \sigma_i^2 \neq \sigma_j^2,\, i, j \in K, i \neq j, \\ &\quad \text{for at least one combination of } i \text{ and } j. \end{aligned}$$

The Bartlett's test statistic is

$$B = \frac{(N-k)\ln\left(\frac{\sum_{i\in K}(n_i-1)\sigma_i^2}{N-K}\right) - \sum_{i\in K}(n_i-1)\ln(\sigma_i^2)}{1 + \frac{1}{3(k-1)}\left(\sum_{i\in K}\frac{1}{1-n_i} - \frac{1}{N-k}\right)},$$

with

$$B \sim \chi^2(k-1),$$

If the test statistic is larger than the critical value, then H_0 is rejected.

Test for autocorrelation

To test for autocorrelation, the Durbin Watson (DW) test is performed on the regression. The test statistic determines whether the serial correlation between residuals of the MLR is clearly present.

Definition 3 (DW test)**.** Consider the set of residuals $\mathbf{e} = \{e_1, e_2, \ldots e_{|T|}\}$ from a OLS regression. Assume that the residuals are stationary (constant mean, variance, and correlation over time), and normally distributed with mean zero and variance σ^2. Define

$$\begin{aligned} H_0 &: \mathbb{E}[e_t e_{t-1}] = 0\, \forall t \in \{2, |T|\} \\ H_a &: \mathbb{E}[e_t e_{t-1}] \neq 0, i \in \{2, |T|\} \\ &\quad \text{for at least one combination of } t \text{ and } t-1. \end{aligned}$$

The DW test statistic is defined as

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2},$$

with

$$0 \leq d \leq 4.$$

The set indicates zero autocorrelation when the test statistic is 2. Then, H_0 is not rejected. Otherwise, the test indicates autocorrelation exists and H_0 is rejected. Positively autocorrelated residuals are found when d ranges between 0 and less than 2 while values of d ranging from above 2 until 4 indicate negative autocorrelation. Since As it is statistically challenging to attain zero autocorrelation, H_0 is not rejected when the test statistic lies within the range of 1.5 and 2.5.

The tests are performed on the experimental results of the regression model and the test statistics are described in Appendix B.1.

3.3 Forecast accuracy

This section elaborates on the forecast accuracy of the regression model. The MAPE will be described. Another measure to examine more precisely is the coefficient of determination, also known as R-squared (or R^2). Since the research deals with promotional data, a categorical R-squared is developed to adjust for the clustered data with different cluster averages.

3.3.1 Mean absolute percentage error

The MAPE is an intuitive measure in terms of relative error and indicates the goodness of fit of the regression formula [10]. It determines the deviation of the forecast in proportion to the actual value of the dependent variable. Then it averages the absolute ratio to retrieve the MAPE. Hence, the MAPE for product i is

$$\text{MAPE}_i = \frac{1}{|T|}\sum_{t\in T}\left|\frac{y_{it}-\hat{y}_{it}}{y_{it}}\right|.$$

The MAPE is always greater than or equal to zero. Moreover, close to zero indicates a small error, thus the regression formula is a good fit for the data. A downside of the MAPE is if the data contains actual values of the dependent variable close to zero, Then, even arbitrary errors are pictured as a huge percental error. Despite that, the MAPE works well on a large data set.

3.3.2 Coefficient of determination

Statistics will provide insight into the correct predictability of the regression model compared to the actual data. Within econometrics, the coefficient of determination is one of the first statistics to assess. The coefficient describes how much the variance of the dependent variable is explained by the independent variables.

Definition 4 (R-Squared)**.** Consider a data set for product i on time horizon T, with actual values $\mathbf{y_i} = \{y_{i1}, y_{i2}, \dots y_{iT}\}$ and predicted values $\hat{\mathbf{y}}_\mathbf{i} = \{\hat{y}_{i2}, \hat{y}_{i3}, \dots \hat{y}_{iT}\}$ obtained from regression formula with OLS estimated parameters. Then, R-squared is calculated by

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\sum\limits_{t\in T}(y_{it}-\hat{y}_{it})^2}{\sum\limits_{t\in T}(y_{it}-\bar{y}_i)^2}, \tag{3.3}$$

where RSS is the sum squared residuals and TSS is the total sum squared.

From Equation (3.3) it becomes clear that the sum squared residual (RSS) of the prediction is presented in proportion to the total sum squared (TSS) where the TSS is the deviation of the real data to its mean. The fraction, RSS/TSS, is the unexplained variance of the prediction. Hence, subtracting it from 1 will return the explained variation as a percentage of the total variation.

The coefficient of determination [7] usually is between 0 and 1, where 1 suggests that the prediction perfectly fits the actual data. In a linear regression without constraints, the R-squared corresponds to the square of the multiple correlation coefficient. If $R^2 = 0$, the fitted line is horizontal and RSS equals TSS. Nevertheless, in some cases, the coefficient of determination can be negative when the prediction outcome is worse than constantly predicting the sample mean. This is the case when

$$\sum_{t \in T} (y_{it} - \hat{y}_{it})^2 > \sum_{t \in T} (y_{it} - \bar{y}_i)^2.$$

Therefore, Equation (3.3) can be smaller than 0. In linear regression models, a negative R-squared is possible when the intercept or slope is restricted, while nonlinear regression models without an intercept can have $R^2 < 0$.

3.3.3 Categorical coefficient of determination

In this report, the coefficient of determination is calculated on categorical data. The data can be split into two clusters, low and high sales volumes. In general, the sales volumes are low in non-promotional weeks and high in promotional weeks. An example of random data is given in Figure (3.1). Each cluster has its own mean value (blue and red line) which significantly differs from the mean of the total data set (green line). Therefore, it is questionable whether the R^2 is a proper method to depict the forecast accuracy.

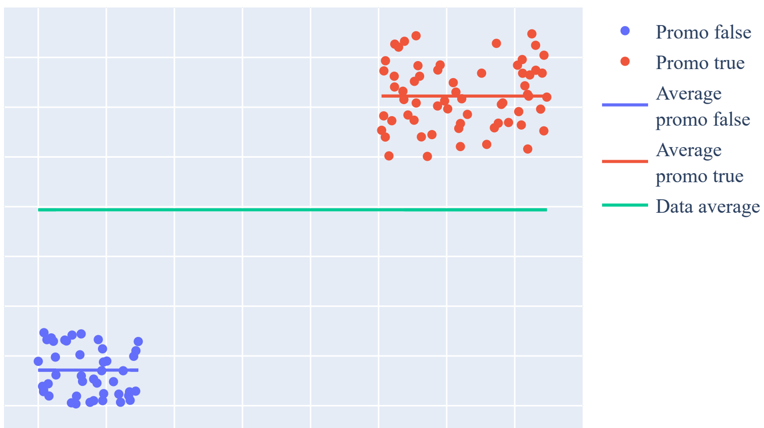


FIGURE 3.1: Random data clusters.

An attempt to improve the correctness of the goodness-of-fit measure includes separately analyzing the forecast accuracy of each cluster,

$$(C_i)_k = \{t \,:\, \text{data point } (i,t) \text{ satisfies cluster condition of } k,\ t \in T\},$$

with $k \in K$, where K is the set of clusters. In this report, the data points are clustered based on the promotional feature resulting in two clusters such that $K = \{0, 1\}$. So, if product i is promoted at time t, then the time point is assigned to cluster $(C_i)_1$, otherwise to cluster $(C_i)_0$. Hence,

$$(C_i)_k = \{t \;:\; x_{it} = k,\; t \in T\}.$$

All clusters are collected in the vector C_i,

$$C_i = \begin{bmatrix}(C_i)_0 & (C_i)_1\end{bmatrix}^T.$$

In this case, the vector C_i has two elements, a non-promotional and promotional cluster. $(C_i)_0$ represents the vector of time points t product i is not promoted, i.e. $x_{it} = 0$, and $(C_i)_1$ represents the vector of time points t product i is promoted, i.e. $x_{it} = 1$. The cluster vectors $(C_i)_k$ are disjoint and the union represents the time set T.

To retrieve the categorical R^2, for all clusters separately, the coefficient of determination is calculated using the cluster mean defined as $(\bar{y}_i)_k$

$$(\bar{y}_i)_k = \frac{1}{|(C_i)_k|} \sum_{t \in (C_i)_k} y_{it}$$

for $k \in K$. This leads to

$$R_k^2 = 1 - \frac{\sum\limits_{t \in (C_i)_k} (y_{it} - \hat{y}_{it})^2}{\sum\limits_{t \in (C_i)_k} (y_{it} - \bar{y}_i^k)^2}.$$

The coefficient of determination of each cluster is combined to retrieve one forecast accuracy. The operations for combining these accuracy's are average and weighted average.

Definition 5 (Categorical R-Squared)**.** Consider the vector C_i representing all clusters and R_k^2 for $k \in K$. Then the average categorical R-squared is

$$R_a^2 = \frac{1}{|K|} \sum_{k \in K} R_k^2. \tag{3.4a}$$

The weighted average categorical R-squared is

$$R_w^2 = \sum_{k \in K} w_k R_k^2 \quad \text{with} \quad \sum_{k \in K} w_k = 1. \tag{3.4b}$$

The weights of (3.4b) are based on the fraction of data points in the cluster compared to the total data points or can be set according to the importance of the clusters. From a marketing perspective, the cluster that is responsible for the highest sales is assumed to be the most important.

The MAPE, R^2, and categorical R^2 are calculated on the results of the regression model and analyzed in Section 5.3.1 to obtain insights into the forecast accuracy of the estimated parameters.

Chapter 4

Promotion calendar optimization by mixed integer programming

This chapter gives an overview of former literature research and presents the mathematical program for optimizing promotional offers. This includes the documentation of the preliminaries and linearization scheme of the nonlinear mixed-integer program (MIP).

4.1 Literature research

Promotional optimization is common practice for manufacturers and retailers of FMCG. A significant part of the marketing budget is dedicated to promotion planning [9]. In the competitive FMCG market, price cuts often occur. Hence, a profitable promotion strategy has become more important in a competitive market [39]. In many stores, promotion planning is performed manually despite the complexity and the large number of products. Therefore, the number of studies on the optimal promotional strategy has increased recently.

Due to new technologies, dynamic pricing has become common practice in the retail sector [11]. Therefore, not only insights into the effect of price changes and price elasticity are desired statistics, but also modern techniques that advise an optimal strategy for pricing and promotion strategies. Subramanian and Sherali [36] developed a multinomial logit nonlinear optimization model. Then, it is transformed into a discrete nonlinear fractional program that maximizes the profit margin while meeting practical constraints. The model controls the observed effects of price elasticity and competition on sales. To retrieve an optimal outcome, the Reformulation-Linearization Technique *(RLT)* is applied to retrieve a MIP such that any commercial optimization software package can solve the linear programming relaxation. The results of the model have selected the price points from a predetermined set that maximizes the profit margin.

Also, the research of Cohen et al. [9] developed a model that selects a price for each time period for one product. The prices to select from are the general price and a range of promotion prices. The general price is chosen, or one of the promotional prices is selected. The objective function of the nonlinear MIP is the profit margin which is determined by the demand function. The demand function is dependent on the price chosen. Moreover, the demand function considers the pre- and post-promotion dip in sales and is expressed in a multiplicative formula of the past prices, seasonality, and trend effects. Business constraints are defined to make sure the problem description matches real-world circumstances, i.e. constraints that follow from negotiations, contracts, or physical and financial limitations. An illustrative example of a business constraint could be the prescribed limitation on the number of promotions that can be executed within a defined time horizon. It is concluded that the model is a good fit. However, this MIP considers only the optimization of the promotion planning of one product, while a manufacturer or retailer has multiple product lines or brands for sale.

The research is extended for multiple items in supermarkets by Cohen et al. [8]. For the promotion optimization problem for multiple FMCG, they have formulated a nonlinear MIP. The objective function and constraints are similar to the model for a single item incorporating cross-time effects. The demand function, however, includes more economic factors, e.g. the post-promotion dip and cross-item effects. Since the demand functions are nonlinear, the problem is solved by integer programming approximations. Nonetheless, for a price set consisting of two values, typically a general price and a promotion price, the problem can be solved efficiently by linear programming. An integer solution is obtained by a rounding method.

From the retailers' and consumers' perspective, Maddah et al. [25] have developed a bi-objective optimization that simultaneously maximizes the producers' and consumers' surplus while deciding on the pricing strategy of the retailer's product line. Once again, a price-dependent demand function is employed in the objective functions. The ε-constraint method is utilized to obtain Pareto optimal solutions. The linearization scheme transforms the formulation into an integer linear program which can be solved for large instances with tens of products and numerous consumer segments.

In the mentioned literature, linearization of the nonlinear optimization problems is a necessary transformation. First, the RLT constructs a set of ranked polyhedral relaxations. These relaxations span the spectrum of convex relaxations, for example, ordinary linear, continuous, and convex hull representations [35]. On the other hand, the linearization of functions can be achieved by piecewise linear approximations. The main idea is to find a range of linear functions that represent local parts of the nonlinear function [14],[22].

4.2 Methodology

For the second part of this research, the goal is to develop a mathematical program that returns a promotion calendar providing information on each product and its promotion weeks. The calendar is optimized by maximizing the sales volume in hectoliters.

4.2.1 Preliminaries

The optimization program requires input from the regression model and additional information. Firstly, the set of time, manufacturers and products are retrieved from Section 3.2.1. The regression formula returning $\hat{y}_{it}$, the predicted log sales volume of product i in week t, is employed in the optimization problem.

The formula is simplified by using the estimated parameters, seasonal effect and median sales of product i at time t. Then, the first part of the forecast formula in Equation (3.1b) is

$$\hat{\alpha}_{it} = \gamma_{1i} log(\pi_t)) + \gamma_{2i} log(\tilde{y}_i). \tag{4.1}$$

Secondly, business constraints have to be defined. For the given time period T, there is a maximum number of points in time a product can be promoted. The variable N_i is the number of promotion slots of a product $i \in I$ on a specified time period T. The preliminary variable vector is

$$N = \begin{bmatrix} N_1 & N_2 & \dots & N_{|P|} \end{bmatrix}$$

Moreover, it is possible to exclude the simultaneous promotion of a pair of cannibalizing products. Therefore, the pr matrix is defined as a symmetric matrix where each entry (i, j) presents if product $i \in I$ and product $j \in J \backslash \{i\}$ can be promoted concurrently. A value of 1 for pr_{ij} indicates that products i and j cannot be promoted simultaneously, while a value of 0 denotes that simultaneous promotion is possible. This implies that the diagonal entries are defined 0.

To optimize the promotion calendar, the decision has to be made when to offer the product for a promotional price. Therefore, the promotion variables of the regression formula are transformed into decision variables in the mathematical problem formulation. Since the meaning remains the same as in Section 3.2.1, the decision variable is an integer variable defined as

$$x_{it} = \begin{cases} 1, & \text{if promoted at time } t \\ 0, & \text{otherwise.} \end{cases}$$

4.2.2 Mathematical problem

Bearing in mind the preliminary and decision variables, the mathematical program can be defined. From the perspective of the manufacturer, the manufacturer wants to maximize the sales volume for all products in its assortment under several business constraints. That is a maximum number of promotion slots per time horizon, and the exclusion of the same promotion times for selected products. This might hold for the manufacturer's own products as well as products from competitors. The latter restriction is a criterion from the retailer.

Within the business domain, the manufacturer that commands the greatest market share is granted first priority to engage in negotiations with the retailer regarding the promotion calendar. Hence, a mathematical optimization program aimed at maximizing the manufacturer's sales volume can be considered effective, provided that it also results in profitable outcomes for the retailer. However, focusing solely on maximizing the manufacturer's sales volume may lead to irrational strategies among competitors. To attain mutually beneficial outcomes, the actions and reactions of competitors should be taken into account. Since the retailer also wants to maximize their profit, the original mathematical program maximizes the sales volume for all products and is defined as

$$\max \quad \sum_{t\in T}\sum_{i\in P} y_{it} \tag{4.2a}$$

$$\text{s.t.} \quad y_{it} = \pi_t \tilde{y}_i \prod_{j\in P} \beta_{1j}^{x_{jt}} \beta_{2j}^{x_{jt-1}} \qquad \forall i \in P, t \in T \tag{4.2b}$$

$$\sum_{t\in T} x_{it} = N_i \qquad \forall i \in P \tag{4.2c}$$

$$x_{it} + x_{jt} \leq 2 - pr_{ij} \tag{4.2d}$$

$$\forall i \in P_{m_1}, j \in P_{m_2}, m_1, m_2 \in M, m_1 \neq m_2, t \in T$$

$$y_{it} \in \mathbb{R}^+ \qquad \forall i \in P, t \in T \tag{4.2e}$$

$$x_{it} \in \{0, 1\} \qquad \forall i \in P, t \in T. \tag{4.2f}$$

In brief, the constraints consist of the definition of the sales volume based on the regression formula described in Chapter 3, the number of promotions, and the conflicting constraint that prohibits simultaneous promotions of products from different brands. A detailed description of these constraints is provided after the final optimization problem formulation in Equation 4.5. It should be noted that the mathematical program presented is a non-linear formulation. Therefore, the problem formulation is linearized in two steps. First, the log transformation is performed on Equation (4.2b) in the same way as in Subsection 3.2.2. The result is the same as in Equation (3.1b) where the first part is replaced by Equation (4.1). Summarizing, Equation (4.2b) is replaced by

$$\hat{y}_{it} = \hat{\alpha}_{it} + \sum_{j\in P} \hat{\beta}_{1j} x_{jt} + \hat{\beta}_{2j} x_{jt-1} \; \forall i \in P, t \in T.$$

Therefore the objective function should be adjusted to make sure the problem remains unchanged, resulting in

$$\sum_{t\in T}\sum_{i\in P} e^{\hat{y}_{it}}. \tag{4.3}$$

The second step in linearizing the problem formulation is introducing linear constraints that approximate the exponential function. The function is split into intervals of the same length where the coordinates of the start and end points are known. A linear function of the form $y = ax + b$ valid on the interval is defined by linear interpolation. This is done for all intervals, such that the complete exponential function is approximated by the method referred to as piecewise linear approximation (PLA). Figure (4.1) shows the exponential function presented in blue. The PLA is shown by the red dashed graph while the red dots present the start and end points of the intervals.

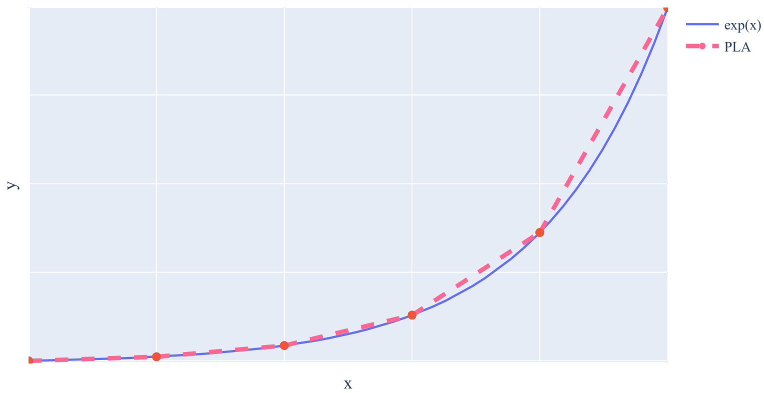


FIGURE 4.1: PLA of exponential function.

The slope a and the intercept b are interval dependent, such that $y = a_k x + b_k$ for $x \in \mathcal{PL}(k)$ with $\mathcal{PL}(k) = [x_k, x_{k+1})$. The focus is to determine the number of intervals to ensure the error, $e^x - y$, is arbitrarily small. While increasing the number of intervals reduces the error, it also leads to a longer computational time. Thus, it is essential to determine the appropriate number of intervals that ensures both a reasonable calculation time and a sufficiently small error. Furthermore, if x is very large, then y is too. Consequently, for x values surpassing a predefined threshold τ, the approximated value y becomes impractically large. To decide what the value of τ is, again the calculation time and error should be minimized. The value of τ is based on the range of x values that require exponential approximation.

Therefore, the constraints defining PLA of the exponential function are

$$\ddot{y}_{it} = a_k \hat{y}_{it} + b_k, \text{ if } \hat{y}_{it} \in \mathcal{PL}(k), \tag{4.4}$$

where $\ddot{y}_{it}$ presents the approximated sales volume in hectoliters. Also, $\hat{y}_{it}$ falls into exactly one interval of $\mathcal{PL}$ ensuring that $\ddot{y}it$ is uniquely defined. In other words, if $\ddot{y}_{it} \in [x_k, x_{k+1}]$, then $\ddot{y}_{it} = a_k \hat{y}_{it} + b_k$.

To define an if-else statement in a MIP, the big-M constraint is employed. The PLA constraint in (4.4) is written as

$$\begin{aligned}
\ddot{y}_{it} - a_k\hat{y}_{it} - b_k &\leq M(1 - z_{itk}) \\
\ddot{y}_{it} - a_k\hat{y}_{it} - b_k &\geq -M(1 - z_{itk}) \\
\sum_{k \in PL} z_{itk} &= 1 \\
z_{itk} &\in \{0, 1\}
\end{aligned}$$

Moreover, the objective function of (4.3) is replaced by

$$\sum_{t \in T} \sum_{i \in P} \ddot{y}_{it}.$$

The final MIP formulation is

$$\begin{aligned}
\max \quad & \sum_{t \in T} \sum_{i \in P} \ddot{y}_{it} && && \text{(4.5a)} \\
\text{s.t.} \quad & \ddot{y}_{it} - a_k\hat{y}_{it} - b_k \leq M(1 - z_{itk}) && \forall i \in P, t \in T, k \in PL && \text{(4.5b)} \\
& \ddot{y}_{it} - a_k\hat{y}_{it} - b_k \geq -M(1 - z_{itk}) && \forall i \in P, t \in T, k \in PL && \text{(4.5c)} \\
& \sum_{k \in PL} z_{itk} = 1 && \forall i \in P, t \in T, k \in PL && \text{(4.5d)} \\
& \hat{y}_{it} = \hat{\alpha}_{it} + \sum_{j \in P} \hat{\beta}_{1j} x_{jt} + \hat{\beta}_{2j} x_{jt-1} && \forall i \in P, t \in T && \text{(4.5e)} \\
& \sum_{t \in T} x_{it} = N_i && \forall i \in P && \text{(4.5f)} \\
& x_{it} + x_{jt} \leq 2 - pr_{ij} && && \text{(4.5g)} \\
& \qquad \forall i \in P_{m_1}, j \in P_{m_2}, m_1, m_2 \in M, m_1 \neq m_2, t \in T && && \\
& \hat{y}_{it}, \ddot{y}_{it} \in \mathbb{R} && \forall i \in P, t \in T && \text{(4.5h)} \\
& x_{it} \in \{0, 1\} && \forall i \in P, t \in T. && \text{(4.5i)} \\
& z_{itk} \in \{0, 1\} && \forall i \in P, t \in T, k \in \mathcal{PL}. && \text{(4.5j)}
\end{aligned}$$

Next, the explanation of formulation is given.

(4.5a) The objective function in (4.5a) is the sum of the predicted sales volume (in hectoliters) for all products over the complete time horizon.

(4.5b) Constraint (4.5b) defines the upper bound for the PLA of $\ddot{y}_{it}$ for $i \in P$, $t \in T$. When $z_{itk} = 1$, then the left-hand side of the equation must be smaller than 0, otherwise it must be smaller than M. $k \in \mathcal{PL}$ indicates which values of a_k and b_k to use.

(4.5c) Constraint (4.5c) defines the lower bound for the PLA of $\ddot{y}_{it}$ for $i \in P$, $t \in T$. When $z_{itk} = 1$, then the left-hand side of the equation must be greater than 0, otherwise it must be greater than $-M$. Together with (4.5b), the lef-handside of the equation is 0, when $z_{itk} = 1$

(4.5d) Constraint (4.5d) ensures that $\ddot{y}_{it}$ for $i \in P$, $t \in T$ is uniquel defined in (4.5b) and (4.5c).

(4.5e) Constraint (4.5e) is the regression formula that defines the log sales volume depending on a preliminary variable, and the promotion variable of all products in set P.

(4.5f) Constraint (4.5f) restricts the total number of promotions of a product within the time horizon T.

(4.5g) Constraint (4.5g) restricts whether products from different brands can be promoted simultaneously at time t. If pr_{ij} is 0, products i and j can be promoted in the same week, if pr_{ij} is 1, at most one product, either i or j or none, can be promoted in week t.

(4.5h) Constraint (4.5h) ensures that the log sales volume and the sales volume are real numbers.

(4.5i) Constraint (4.5i) ensures that the promotion variable of product i at time t, x_{it} is a binary variable.

(4.5j) Constraint (4.5j) ensures that the big-M variable of product i at time t for interval k is a binary variable.

Chapter 5

Results

This chapter begins with a description and analysis of the data for the case study. The prescribed format of the data for the regression model and optimization problem is provided, along with an explanation of how it is extracted from the data. Thereafter, the empirical results of the regression model and promotion calendar optimization problem are presented.

5.1 Data

To validate the regression model and promotion optimization calendar described in Chapter 3 and 4, a transaction data set of 2020 is utilized. However, both methods require input data in a specific format. Despite not all data of the variables is directly observable from the transaction data, assumptions are made to attain all information. In this section, the data is analyzed and the assumptions are elucidated.

The data set contains weekly transaction data of 2020 of one retailer chain in the Netherlands. Each data point represents the sales of a specific product, in a specific week. Weekly data includes the number of units sold and the revenue, while the master data of the products contain information on the manufacturer, brand, and content in milliliters. The product name is referred to as upc which stands for unique product code. A sample row of transaction data is shown in Table (5.1).

upc	manufacturer	brand	content	week	store	units	sales

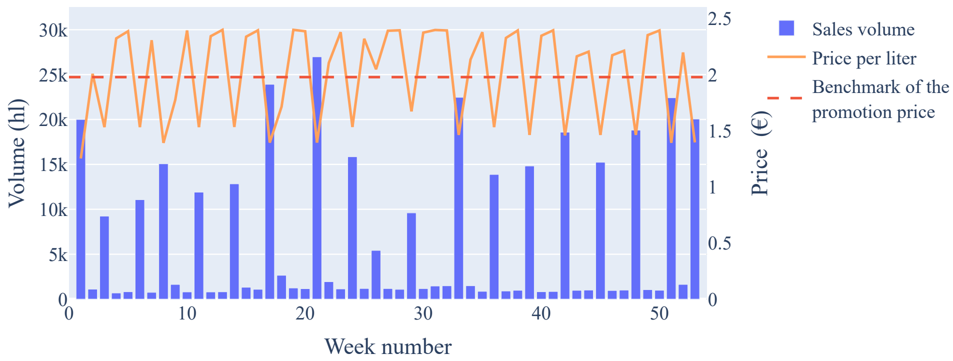
Table 5.1: Example of the data.

The data set includes 1242 products offered by 22 manufacturers, which are available at promotional prices at least once annually. Furthermore, it is necessary to know which products cannibalize each other to define product groups. This research focuses on a single product group, which consists of the most sold product of the 3 largest manufacturers shown in Table (5.2). This table also shows the cumulative market share of the products and the manufacturers of the total market. The largest 3 manufacturers are responsible for 80% of the sales within the market, and the market share of manufacturers 1, 2, and 3 are respectively 28%, 39%, and 13%. In terms of revenue, the gap between the largest and third largest manufacturer is already 26%. Similarly, it can be observed that the market share of the top three most sold products, *upc* 1, 2, and 3, are 11%, 9%, and 6%, respectively. This suggests a gradual decrease in market share among the products. Nevertheless, a cumulative market share of 26% for the 3 products is significantly large, which suggests they are each other biggest competitors.

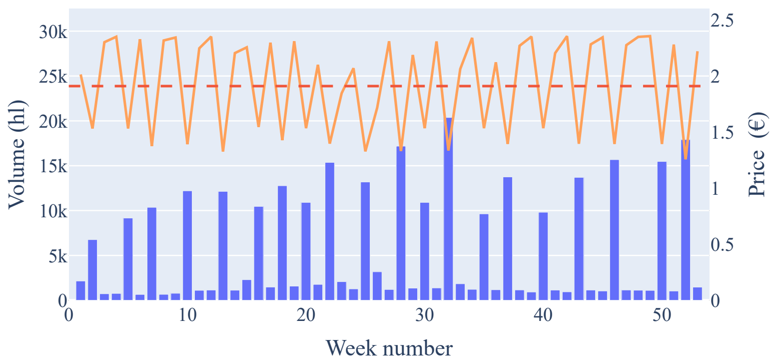
product	**manufacturer**	**product share**	**manufacturer share**
upc 1	*manufacturer* 1	11%	28%
upc 2	*manufacturer* 2	9%	40%
upc 3	*manufacturer* 3	6%	13%

TABLE 5.2: Product manufacturer combination and market share of total market revenue.

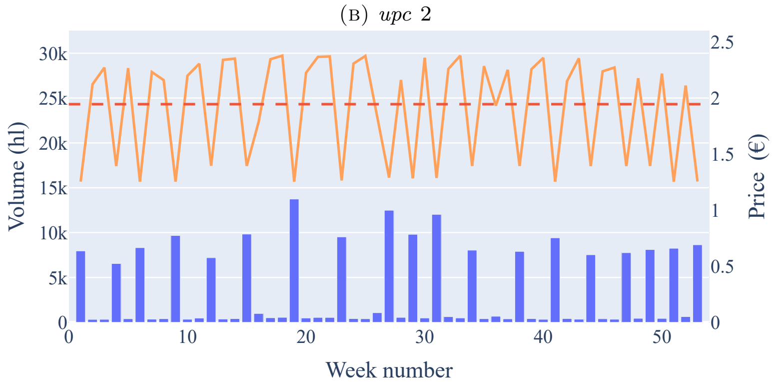
The regression model from Chapter 3 requires the actual sales volume, the median sales volume, and the promotional transaction status for each product group. Since the store-level data is not of interest in the regression model, the sales data in Table (5.1) is aggregated resulting in national sales transaction data per week. Directly from the aggregated data, the actual sales volume is obtained and the median sales volume can be calculated. To gain insight into the promotional transaction status, it is necessary to establish a benchmark for the promotion price, which can determine if the product was promoted during the given week. The benchmark of the promotion price is established at 90% of the median price of the product. The median price represents the lower bound of the base price for which the product is sold the majority of the weeks. Since it is assumed that the number of promotion weeks is less than 40% of the 53 weeks in a year, the median price is slightly too high as an upper bound for the promotion price. Therefore, there is chosen to decrease the benchmark to 90% of the median price. In Figure (5.1) the price per liter is shown across time as well as the benchmark of the promotion price. Each price point under the dashed red graph presents a promotional week of the product resulting in 20, 22, and 22 promotion weeks for *upc* 1, 2, and 3 respectively. Besides the liter price, the sales volume in hectoliters is shown in the figure. It can be observed that sales are high when the price is low and an incremental in the sales volume indicates a promotion week. For *upc* 1, in Figure (5.1a), weeks 2 and 26 have a price just above the benchmark of the promotion price. Although in week 2 the price is very close to the benchmark, the price is higher than in week 1 and week 3 and the sales do not show an incremental in comparison to other non-promotional weeks. On the other hand, the sales in week 26 show an increase, and the price in the week before and after have a higher price than week 26. Remarkable in the promotion determination of *upc* 2 is in week 23, which is counted as a promotion, but does not have an inflated sales volume. In this case, the price was lower the week prior, and therefore the effect of the promotion did not lead to a significant increase in the sales volume. However, the sales volume barely differs from other weeks with a non-promotional price due to which it is questionable to consider this week as a promotional week. Lastly, the price of *upc* 3 in week 36 is considered a promotion price. Although the prices in the week before and after are higher, the sales volume is very similar which makes it questionable if this is an actual promotion week. All in all, the benchmark of the promotion price is not robust in finding all promotion weeks and misses or finds additional promotion weeks. A small increase or decrease in the benchmark might not only remove wrongly determined (non-) promotions but also attain new misjudged points.



(A) *upc* 1



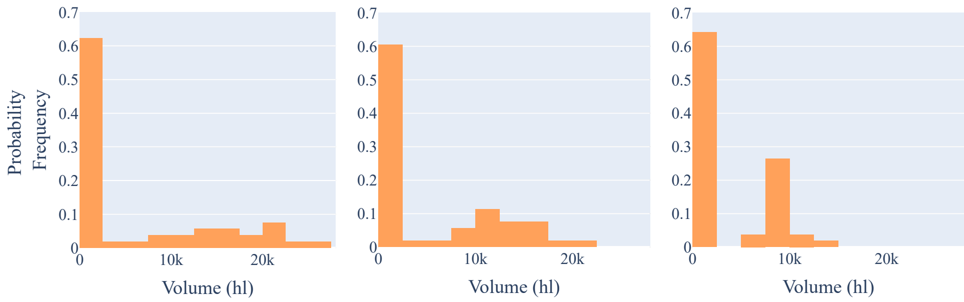
(B) *upc* 2



(C) *upc* 3

FIGURE 5.1: Price per liter, sales volume in hectoliter and the benchmark of promotion price per product. It can be seen that low prices coincide with large sales volume which indicates a promotional week.

It has been observed that low prices and large sales volumes are commonly occurring events which suggests that non-promotional weeks have small sales volumes. Figure (5.2) shows that the volumes up to 2500 hl, also referred to as base volume, occur vastly 60% of the time for all 3 products. Nevertheless, the large volumes occurring in promotional weeks can be up to respectively 11, 10, and 6 times the base volume of *upc* 1, 2, and 3. On yearly bases, the sales volume achieved by promotion pricing proved to be 88, 87, and 93 % of the turnover in 2020 respectively for *upc* 1, 2, and 3. Hence, the impact promotion pricing has on the sales volume of manufacturers is significantly large. Furthermore, the sales volume in promotional weeks differs enormously per product. Most likely, *upc* 1 and 2 have a stronger nationwide brand recognition than *upc* 3 since the promotion prices of *upc* 1 and 2 are not more extreme than *upc* 3 which becomes clear by comparing the depth of the promotion prices in Figure (5.1). Additionally, the lower base sales volumes of *upc* 3 suggest that the product brand is less well-known.



(A) *upc* 1 (B) *upc* 2 (C) *upc* 3

FIGURE 5.2: Distribution of weekly sales volume which indicates a frequent weekly sales volume up to 2500 hl and less often a large sales volume.

Lastly, Figure (5.3) presents the grouped sales per week which shows a seasonal trend. Mainly in the last weeks of the year, the December month, sales are higher than the rest of the year. Due to the definition of the week numbers, sales in the 1^{st} week are above average as well. Some years have 52 weeks, while others have 53 weeks. Therefore, New Year's Eve either falls in the 53^{rd} week or the 1^{st} week. This dataset includes the sales data from New Year's Eve of 2019 in the 1^{st} week and from 2020 in the 53^{rd} week, which is reflected in the increased sales volume. The incremental sales in weeks 17 and 21 are the effect of Kingsday and Ascension day weekend. Conversely, the period with the lowest sales volume is the month of the year due to 'Dry January'. However, the 1^{st} week of February, week 6, shows a peak in sales that is higher than the Carnaval peak in week 8. The end of 'Dry January' can cause an increase, but the manufacturers do use the opportunity to have their products promoted in the 1^{st} week of February and two manufacturers introduced a promotion price in week 6. This is the last week of June, so consumers who are leaving the country for summer vacation do not stimulate sales either. Although promotions can be profitable, two manufacturers promoting their products in the same week may not always be profitable which is shown by the drop in sales in week

26. Generally, the seasonal effect depends on holidays and festivities, but definitely of influence are the promoting strategies of products. Multiple promotions and festivities in a week reinforce sales increase, while the opposite is true without any festivities.

FIGURE 5.3: Total sales volume of 3 products.

5.2 Data preparation

This section defines the input data for the regression model and subsequently for the promotion calendar optimization problem. The part about the regression model includes an explanation of the time series cross-validation applied.

To develop a regression model, the input variables including seasonal effect, median sales, a promotion variable, and a lagged promotion variable need to be collected for each point in time, and for each product within the product group. The seasonal effect is computed as an index number based on the grouped sales volume presented in Figure (5.3). The median grouped sales are set as the base index whereafter the seasonal index for each week is computed by comparing the weekly grouped sales volume to the base volume. This information is used to estimate the parameters of 3.1b by OLS. To validate the model, usually, one train-test split is performed where 80% of the data is used for the train set and 20% for the test set. However, this research applies a time series cross-validation to evaluate the model in a more robust way. Also, it analyzes the model's capability to make accurate predictions for future time periods. In complex regression models, train data can be fitted perfectly while the performance on unseen data is poor. This is called overfitting. The unseen test data is used to assess the degree of overfitting. Multiple train-test splits are performed to attain insights into the generalizability of the model since it should produce accurate forecasts for a broad range of unseen data. Although this research contains data from one year only, there are 3 train-test splits performed. Each fold estimates the parameters on the train set and the test set is predicted. The test set of Fold 2 is used for the analysis of the results in the next section. Hence, the test set consists of weeks 41 to 53. In the rest of the report, weeks 1 up to 40 is considered as the train set.

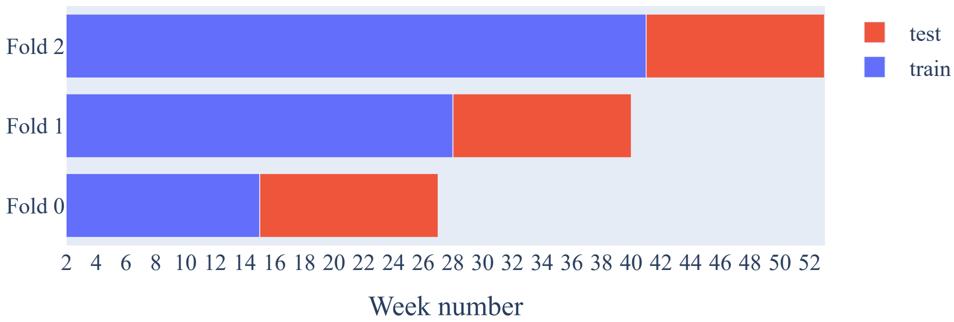


FIGURE 5.4: Train-test split for time series cross-validation, where the week on the left side of the bar is included in the set, and the week on the right side of the bar is not included in the set. e.g. Fold 2, week 41 is excluded in the train set and included in the test set.

The promotion calendar optimization problem uses the regression output as input such that the sales volume is optimized while deciding when the products should be promoted. However, there are several restrictions when planning the promotion of products. First, the number of promotions of a product within a time period is restricted. Therefore, the time period should be known which consists of weeks 41 to 53 since the calendar is optimized for the test set. Then, the optimal predicted value can be compared with the predicted value of the actual promotion calendar. From the data, it is found that *upc* 1, *upc* 2, and *upc* 3 are promoted respectively 5, 4, and 6 times in a 13-week period. This means that for at least 2 weeks, there are 2 products promoted at the same time. In this case, *upc* 1 and *upc* 3 are promoted simultaneously in weeks 51 and 53. Consequently, it should be possible for these products to be promoted at the same time. The input of the pr matrix and N vector indicating the restrictions of product promotions and the number of promotions are given in Table (5.3a).

Despite retrieving the preliminary input for the constraints from data, business decisions can overrule the data. In this market, it is not uncommon that the retailer restricts shared promotion slots for the largest products in the market. Therefore a second situation is defined that forbids promotions for *upc* 1, *upc* 2, and *upc* 3 at the same time slot. Then there are 13 promotion slots to divide across 3 products. The number of promotions should be at most the number of time slots to divide, thus *upc* 1 and 3 have one promotional week less than obtained in the data. The preliminary input data of the constraints for Situation 2 is presented in Table (5.3b).

	upc **1**	*upc* **2**	*upc* **3**	**Number promotions**
upc **1**	0			5
upc **2**	1	0		4
upc **3**	0	1	0	6

(A) Situation 1 retrieved from historical data

	upc **1**	*upc* **2**	*upc* **3**	**Number promotions**
upc **1**	0			4
upc **2**	1	0		4
upc **3**	1	1	0	5

(B) Situation 2 overruled by business decisions

TABLE 5.3: Constraints for promotion calendar optimization of two situations.

The goal of the promotion calendar optimization problem is to return a calendar that maximizes the total sales volume of the products combined accounting for the cannibalization effects. Therefore, the actual calendar is presented in Table (5.4) such that this can be used in the comparison to the later presented optimal calendars. The actual calendar shows two promoting patterns. In weeks 42 and 45, the calendar starts with the promotion of the largest product and continues with the second largest and the smallest, respectively *upc* 1, 2, and 3. The other pattern which begins in weeks 49 and 51 kicks off with the least sold product to continue with the second largest and largest product. Moreover, an overlap in promotion weeks occurs.

week no.	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**
upc **1**													
upc **2**													
upc **3**													

TABLE 5.4: Actual promotion calendar of test set where the red cells are promotion weeks.

5.3 Experimental results

In this section, the prediction performance of the regression model is assessed by employing the input data specified in the previous subsection. This section is split into two parts. The first part presents the results of the regression model where the predicted volumes are compared against the actual volumes whereafter forecast accuracy is evaluated. In the second part, an optimal promotion calendar is solved adjusted for cannibalization effects which are estimated in the regression formula that is used in the optimization problem. For the two problem scenarios, the predicted sales volumes of the optimal calendars are compared with the predicted sales volumes of the actual promotion calendar.

5.3.1 Regression model

The regression model is executed for the product group and returns the estimated parameters of Equation (3.1b) for each product. The coefficients are presented in the appendix Table (B.1) and the time series plot of the actual data and the prediction of the test set for each product is shown in Figure (5.5). The coefficients and their p-values presented in Table (B.1) are fitted on the train set of Fold 2. The coefficients corresponding to the promotion variables of the products, except for the predicted product, reflect the cannibalization effect caused by promotion. The majority of the coefficients have significant p-values, suggesting there is no clear evidence to reject the null hypothesis which indicates that the variable is correlated with the dependent variable and that the relationship between the regressor and thus it is likely that the dependent variable is not driven by randomness. The p-value of the seasonal effect, the median sales, and the promotion variable of the predicted product are smaller than 0.01. Among the 8 regressors, *upc* 1, 2, and 3 have 5, 6, and 5 significant coefficients considering the significance level of 0.1. Although significant coefficients do not necessarily imply a good fit regression, there is enough confidence that the independent variable has a significant impact on the outcome of the dependent variables and thus increases the reliability of the model's prediction.

As pointed out in the previous subsection, the cross-validation is performed using 3 folds. To evaluate the model's accuracy and identify overfitting, the coefficient of determination is calculated on the train and test set of each fold which are presented in Table (5.5). The R^2 on the train set for all products indicates satisfactory accuracy and explains the variance in the model since they are close to 1. An increase in the number of data points of the train set shows a lower R^2 on Fold 1 and Fold 2. Still, around 90% of the variance is explained which indicates that the estimated parameters are a good fit for the data. Nevertheless, it is evident that the R^2 of the test sets collapses which indicates that the model's ability to explain the variance in the data is not as elevated in the train set. For Fold 0, the enormous drop is explained by the number of data points in the train set. In fact, the number of data points in the test set is one less than in the train set, therefore the OLS is applied on too little information resulting in inaccurate estimates. Although the R^2 of Fold 1 and 2 slightly drop for *upc* 1 and 2, there are no clear indications that the regression model fits the noise in the data rather than the fundamental data pattern. Indeed, the R^2 of *upc* 3 increases which signify an accurate fit of the fundamental data pattern. Hence, solely based on the R^2 of Fold 2, the parameters are a good fit for the regression model and return accurate predictions on unseen data.

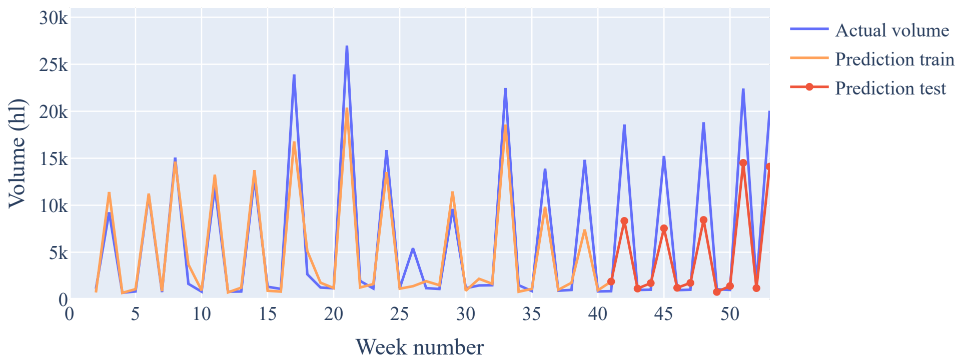
	Fold 0		**Fold 1**		**Fold 2**	
	train	*test*	*train*	*test*	*train*	*test*
***upc* 1**	0.9896	0.5267	0.8995	0.8187	0.8903	0.8583
***upc* 2**	0.9822	0.5357	0.9594	0.8952	0.9500	0.9031
***upc* 3**	0.9977	0.6703	0.9042	0.7783	0.8743	0.9671

TABLE 5.5: Coefficients of determination on the train and test set for each fold in the time series cross-validation.

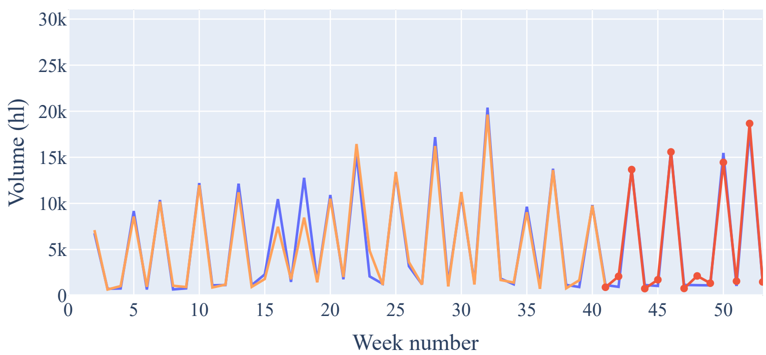
Next, the observations on the coefficient of determination are considered while concluding the visualization of the regression model. Figure (5.5) presents the graph of the actual volume, the fit of the train set, and the predicted volumes by the regression model for all products. Figure (5.5a) represents the output of *upc* 1 revealing that the predictions of weeks 42, 45, and 48 are substantially lower than the actual volumes. Note that those weeks have the exact same input variables resulting in a similar lower prediction volume. Given that the only input variable that differs is the seasonal effect, the small difference between the predicted volumes of the weeks is explained. In the train set, the singular week with the same promotion input data is week 39. Neglecting the effect of the seasonal variable, the predicted value is also substantially lower than the actual volume, with an error of 10.000 hectoliters.

For *upc* 2 in Figure (5.5b), the predictions for the promotion weeks with high sales volumes are accurate. The MAPE, explained in Section 3.3, is 2.4% and 4% for the train and test set respectively. This means that the prediction of the test set has an average deviation of 4% from the actual volume. In contrast with *upc* 1, the promotional sales are not underestimated, because the train set includes multiple weeks with the same combination of promotional input data. One combination in terms of promotional input variables has 6 data points in the train set that predict promotion peaks of weeks 43 and 46 very accurately with an absolute deviation of 0.2% and 0.6% respectively. Therefore, the promotional weeks in the test set react similarly to those in the train set adjusted for the seasonal effect.

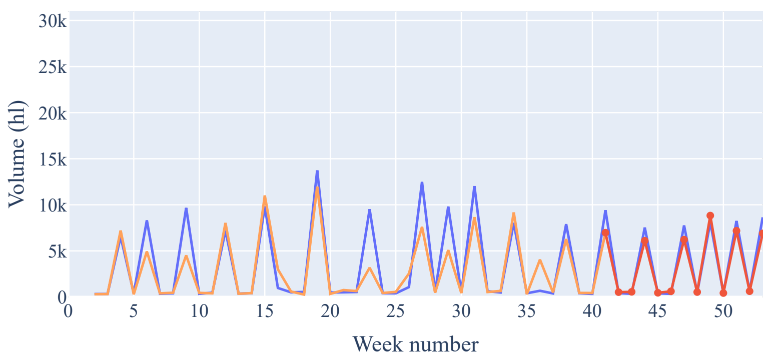
The MAPE of *upc* 3 is 3.1% on the test set which is the lowest MAPE of the 3 products. Furthermore, it scores the highest coefficient of determination and is closer to 1 than the goodness-of-fit measure on the train set. From Figure (5.5c) it can be observed that the predictions of the test set in red fit closer to the actual volumes than the fitted train data points in orange. Therefore, it can be concluded that the regression model for *upc* 3 is the best fit and the most influenced by the promotion strategy of *upc* 1 and 2, rather than *upc* 3 affecting the sales of *upc* 1 and 2 with its promotion strategy. However, an extraordinary data prediction can be observed in week 36 where the sales prediction is larger than the actual volume. This data point has an actual sales volume around the base of non-promo weeks, while it is assigned a promotion week based on the liter price. In principle, a sales increase is expected but this data point is not consistent with the fundamental data pattern. It shows that the weak robustness of the benchmark of the promotion price, as discussed in Section 5.1, can potentially have a negative impact on the regression model's fit.



(A) *upc* 1



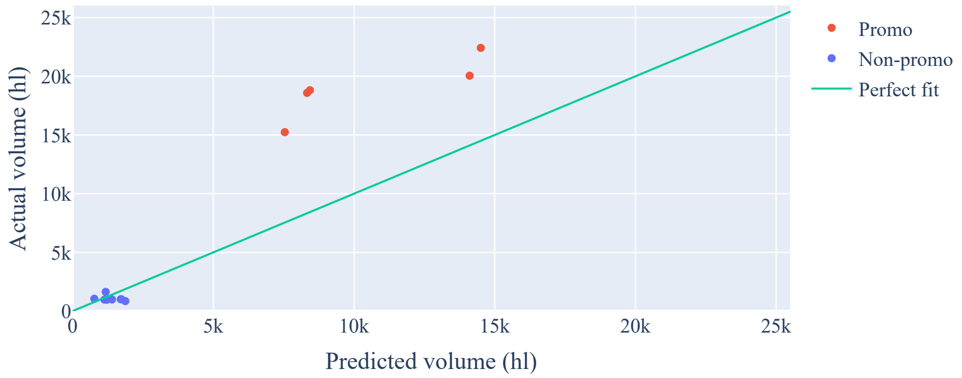
(B) *upc* 2



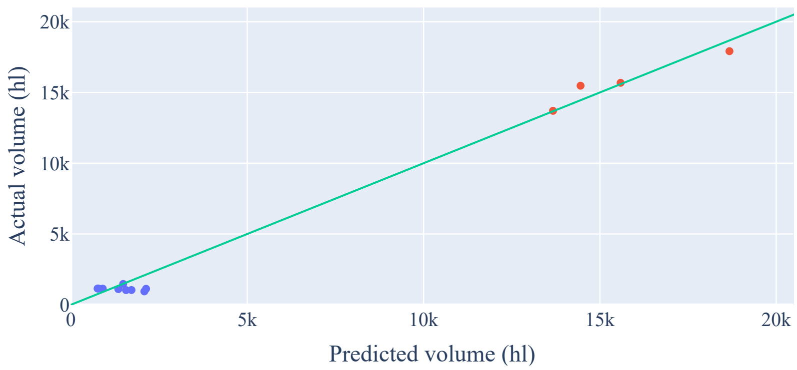
(C) *upc* 3

FIGURE 5.5: Actual and predicted sales volume shown as time series.

Up until now, the MAPE of the regression is calculated on the log values of the prediction instead of the absolute sales volumes in hectoliters, since the regression minimizes the error of actual and predicted log sales volume. When the MAPE is calculated with ν_{it} and $\exp(\hat{y}_{it})$, the actual volume and predicted volume in hectoliters, the MAPE of the test set determined on hectoliters is 48%, 34%, and 24% for *upc* 1, 2, and 3 respectively. In comparison to the MAPE on the log values of *upc* 1,2, and 3, that is 6%, 4%, and 3%, the gap is significantly large. For this reason, the forecast accuracy R^2 of the regression model has to be revised as well.

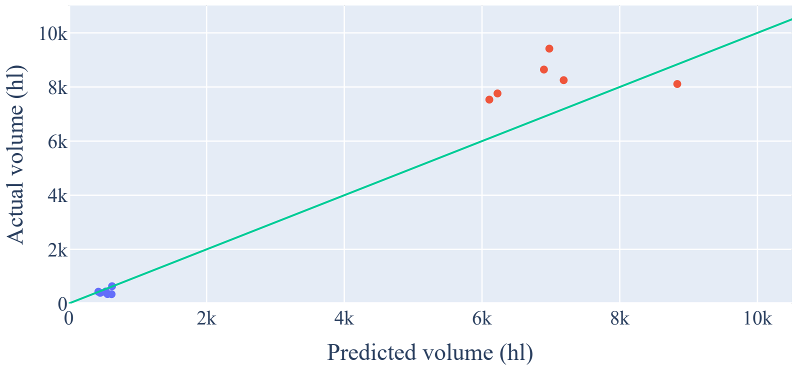


(A) *upc* 1



(B) *upc* 2

FIGURE 5.6: Actual v.s. predicted volumes of test set (a-b).



(c) *upc* 3

Figure 5.6: Actual v.s. predicted volumes of the test set (c).

Firstly, the actuals and predictions per product can be observed in Figure (5.6). Additionally, the perfect fit line is plotted highlighting the data points where actuals are equal to the predictions. The points in the graph being close to the perfect fit, have a small error and indicate a high level of accuracy in predicting the dependent variable. It is observed that the predicted volumes of promotional weeks for *upc* 1 are underestimated and are greatly distanced from the perfect fit corresponding to a lower accuracy of the prediction. This is in accordance with the observations in Figure (5.5a) as well as the conclusions that the forecast accuracy R^2 has to be revised. Also, the promotional predictions of *upc* 3 are underestimated, but not as severely as *upc* 1. Contrarily, the promotional weeks of *upc* 2 are fitted almost perfectly and sufficiently accurately to be used in practice. Despite this, non-promo sales are overestimated in general and less useful in real-world circumstances.

Although several predicted data points are significantly distanced from the actual volume, the R^2 for Fold 2 of the regression remains close to 1 which is observed in Table (5.5). In Section 3.3, it is explained that the error of the predicted and actual volume is compared with the deviation of the actual volume from the mean. Figure (5.7) shows the predicted volume against the actual volume of the train set. The majority of the data points exhibit a smaller deviation from the perfect fit line compared to the mean, hence an R^2 of the train set for this regression of 0.95.

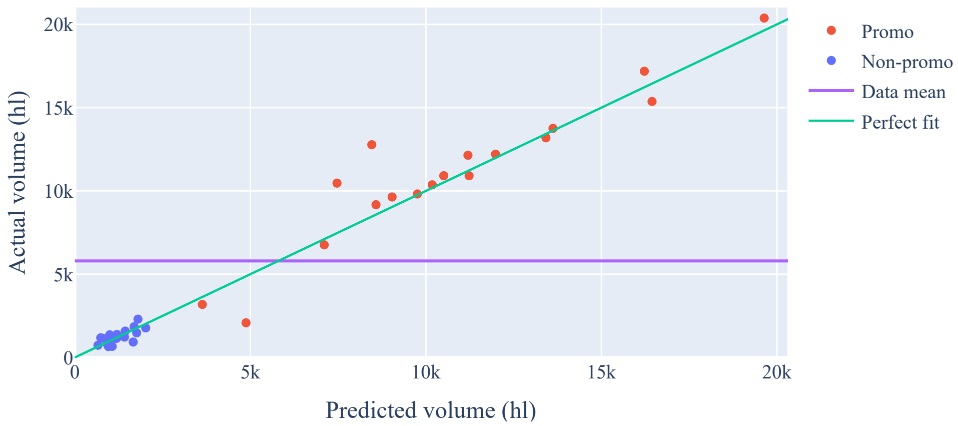
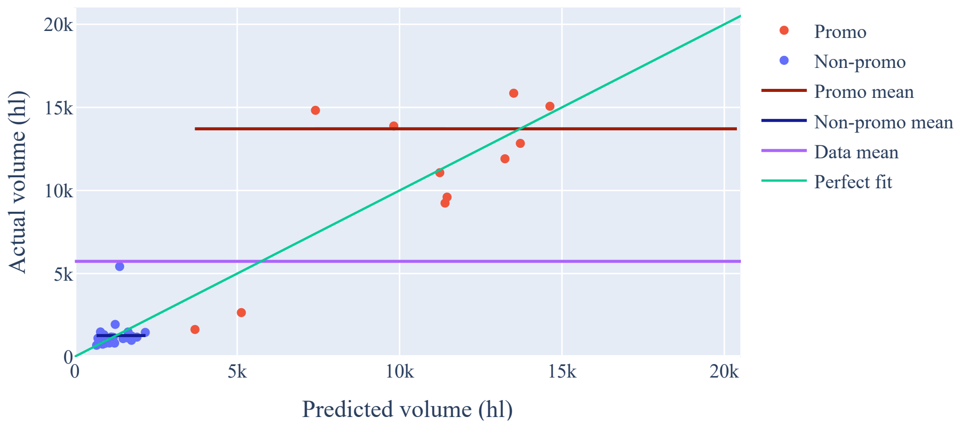


FIGURE 5.7: R^2 visualization of *upc* 2 of the train set.

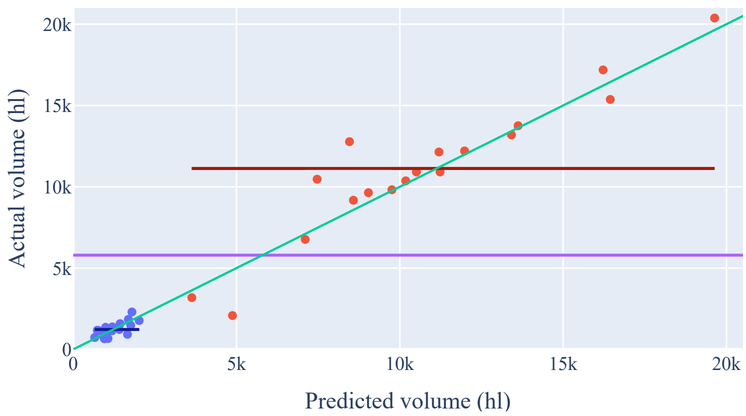
Section 3.3 developed a categorical R^2 such that the predicted volumes can be compared with the cluster mean rather than the data mean. In addition to Figure (5.7), the cluster means of the promo and non-promo data clusters for each product retrieved from the train set are displayed in Figure (5.8). For the categorical R^2, the same comparison is performed for the R^2, only here the focus is to compare the distances of data points to the perfect fit line with the distance to the cluster mean. For both *upc* 1 and *upc* 2, the majority of the promotional data points are found to be closer to the perfect fit line than to the cluster mean indicating a positive R^2 and reaching towards 1 for a declining deviation. The promo data points that are located closer to the non-promo cluster mean than the promo mean are of particular interest. These data points are assigned to the promo cluster based on the liter price, while the volume shows the trend of non-promo data. Nevertheless, the predicted value is near the actual volume, hence the fraction of unexplained variance remains small and the data points are estimated accurately. On the other hand, *upc* 3 has both under and overestimated predictions in the promotional clusters where the underestimated data points lay closer to the cluster mean and overestimated data points are closer to the perfect fit. Compared to the other products, the linear regression coefficients of *upc* 3 resulted in the lowest R^2 of 0.30. The R^2 of *upc* 2 is 0.8 and the model performs the best for this product since the R^2 of *upc* 3 is 0.67. All categorical coefficients of determination for non-promo and promo clusters are presented in the first two columns of Table (5.6).

	non promo	**promo**	**average**	**weighted average of data points**	**weighted average of volume**
***upc* 1**	-0.43	0.67	0.12	-0.03	0.51
***upc* 2**	-0.24	0.80	0.28	0.24	0.68
***upc* 3**	-0.16	0.30	0.07	0.01	0.26

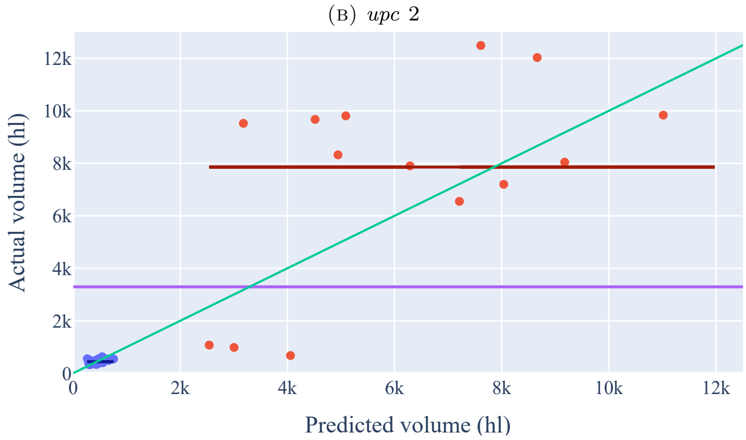
TABLE 5.6: Categorical R^2 and its average, and weighted average on the number of data points and the fraction of volume in cluster.



(A) *upc* 1



(B) *upc* 2

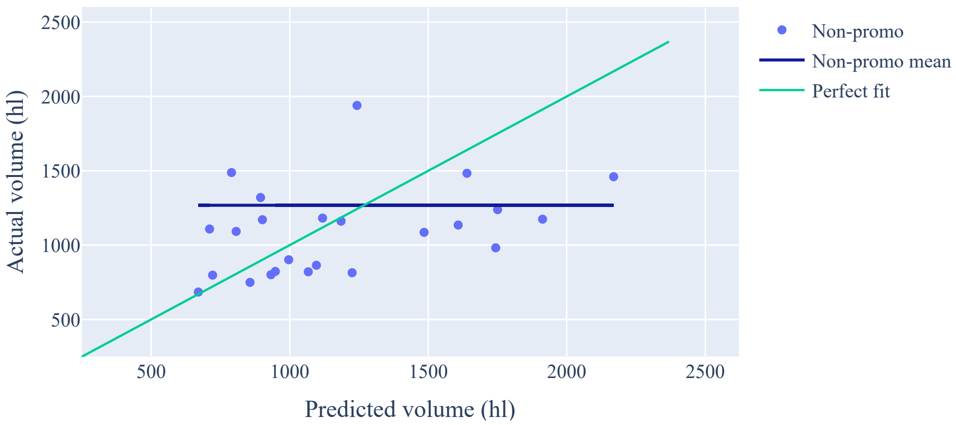


(C) *upc* 3

FIGURE 5.8: Categorical R^2 visualization of the train set.

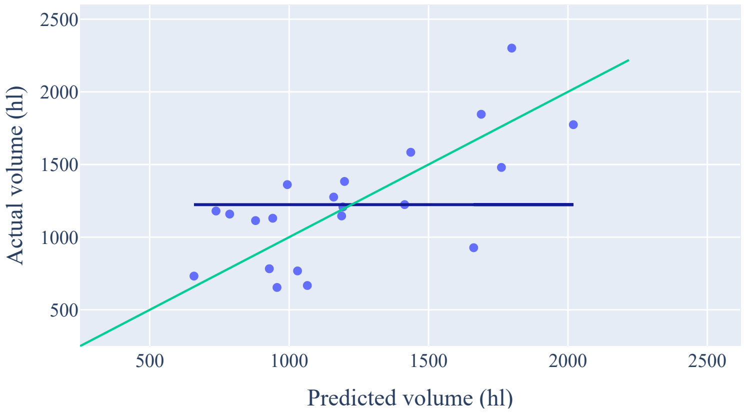
Next, the categorical R^2 of the non-promo cluster is evaluated. Since the volumes are less spread, Figure (B.1) displays crops of the non-promo data points from Figure (5.8). For all products, the data points are situated around the cluster mean with a deviation of circa 750, 750, and 150 hectoliters for *upc* 1, 2, and 3 respectively. Although it is difficult to see, the majority of the data points for each product lay further from the perfect fit than the cluster mean. This means the R^2 of the non-promo clusters are below 0 which is supported by the second column of Table (5.6). The least inadequate model performance is for *upc* 3, but since R^2 is below 0, the cluster mean is a better predictor than the linear regression.

Having the cluster coefficients of determination, there should be one forecast accuracy of the regression. Therefore, the R^2 of both clusters are combined by taking the average or weighted average. When taking the average, each cluster contributes equally to the forecast accuracy measure, despite the fact that the non-promo cluster, which contains more data points than the promo cluster, has a larger impact on the result. To obtain a categorical R^2 that provides insight into the predicted number of values, the weights of the weighted average categorical R^2 are adopted as the size of the cluster. In the fourth column of Table (5.6) it can be observed that *upc* 2 has the highest number of data points with a prediction closest to the actual volume. An alternative approach to determining weights is by considering the cluster that contributes the most sales volume. Since the sales volumes are primarily promotion driven (see Section 5.1), weights are defined based on the volume share of a cluster providing insights into the accuracy of the total predicted sales volume. The fifth column of Table (5.6) shows that the sales volume of *upc* 2 has the highest explained variance.

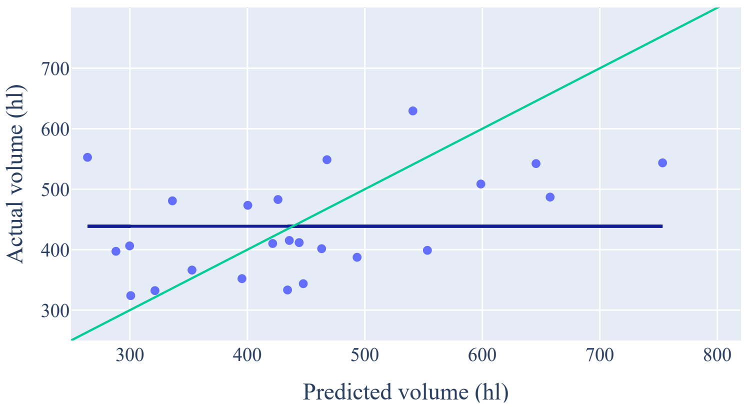


(A) *upc* 1

FIGURE 5.9: Categorical R^2 visualization, zoom of the non-promo cluster (a).



(B) *upc* 2



(C) *upc* 3

FIGURE 5.9: Categorical R^2 visualization of the train set, a close-up of the non-promo cluster (b-c).

5.3.2 Promotion calendar optimization problem

Given that the expected sales volume of the promotion calendar can be predicted by the regression model, the promotion calendar optimization problem is executed simultaneously for all products in the product group utilizing the regression model. The goal is to return a promotion calendar that maximizes the sales volume from the retailer's perspective taking into account the cannibalization effects. However, the preference is that the manufacturers have an increased sales volume compared to the actual calendar. Specifically, the manufacturer that is running the optimization program would prefer to see a higher sales volume for their products. The two scenarios proposed in Table (5.3) were implemented on the test set of the regression model, and their performance was analyzed sequentially. It is important to note that the optimization problem makes decisions on 13 promotion weeks while optimizing the sales volume of the last 12 weeks since the regression model uses lagged variables. Therefore, the sales volume of the 1^{st} week (41) is excluded because the input of the week before should be known to retrieve the sales volume. Although this information can be attained from the train set, fixing this input can affect the promotion strategy the week after. Because the optimal promotion calendar should not be influenced by actual factors, the sales volume of the 1^{st} week is excluded such that the promotion decision in week 41 is primarily based on optimizing the sales volume of week 42. Lastly, the outcomes of the scenarios are compared.

Scenario 1

In Table (5.3a), Scenario 1 suggests having 5, 4, and 6 promotional weeks for *upc* 1, 2, and 3 respectively. In this scenario, it is only possible to promote *upc* 1 and *upc* 3 simultaneously. The result of the optimization problem is a calendar for all 3 products, shown in Table (5.7). Since the volumes in week 41 are not considered in the optimization problem, it is convenient for the model to promote two products in the 1^{st} week while in the actual calendar, the overlapping promotions are scheduled toward the end of the year. End-of-year sales are in general higher due to the holiday season, but regardless two promotions in one week are not found optimal. It is noteworthy to observe the successive promotions of *upc* 3 in weeks 41 and 42. According to the regression model, the effect of a promotion in a preceding week is negatively estimated by the regression model. Although the sales in week 42 are lower than the sales would be without a consecutive promotion, it does not have a negative impact on the total sales of the retailer. Furthermore, a promotion cycle can be observed. That is, after a promotion week of *upc* 3, *upc* 2 is promoted followed by *upc* 1. Nevertheless, some overlap of promotion weeks between products occurs due to the total number of promotions (15) which is more than the time horizon offers (13).

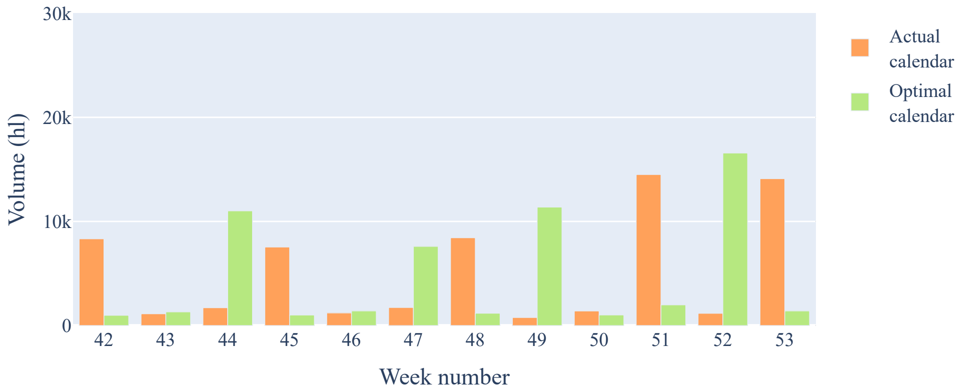
week no.	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**
***upc* 1**													
***upc* 2**													
***upc* 3**													

TABLE 5.7: Optimal promotion calendar in Scenario 1 where the red cells are promotion weeks.

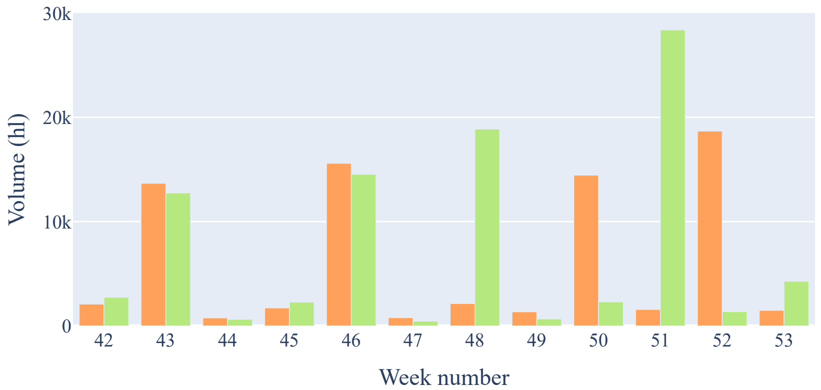
Certainly of interest are the maximized sales volumes of the optimized promotion calendar. Figure (5.10) presents the predicted sales volume of the actual and optimal calendar per week utilizing the estimated parameters of the regression model described in Section 5.3.1. The larger bars are the promotional weeks, which correspond with the promotion calendar in Table (5.4) and (5.7). Since there has been a switch in the promotion weeks of the actual and optimal calendar, the heights of the bars are alternate per week for the predicted and optimal outcome.

The promotion for *upc* 2 in week 51 is clearly more profitable than the actual promotion in week 52, because promoting *upc* 3 instead of *upc* 1 in the preceding week resulted in a weaker cannibalization effect. Even multiple non-promotional weeks resulted in higher sales in the optimal promotion calendar due to promotions of another product in the preceding week resulting in weaker cannibalization effects. It is important to note that for *upc* 3 the number of promotions to compare is equal for the actual and optimal promotion calendar since both have a promotion in week

41. Moreover, it is observed that the sales of 4 promotional weeks in the optimal calendar result in higher sales than 4 promotions of the actual calendar. Only the sales of promotional week 47 are lower in the optimal calendar compared to the actual promotion in the same week. The weekly predicted sales achieved during promotions by *upc* 1 decreased because the predicted sales volume of promotion in week 41 cannot be determined properly resulting in a calendar comparison of 5 actual versus 4 optimal promotions.

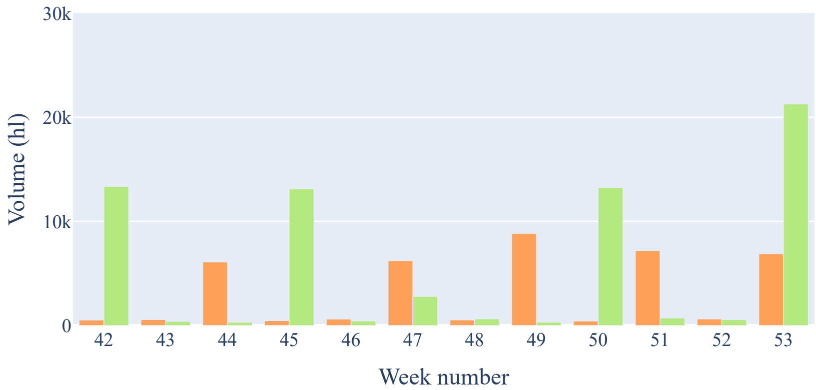


(A) *upc* 1



(B) *upc* 2

FIGURE 5.10: Predicted volumes of actual promotion calendar and predicted volumes of optimal promotion calendar in Scenario 1 (a-b).



(c) *upc* 3

FIGURE 5.10: Predicted weekly sales volumes of actual promotion calendar and optimal promotion calendar in Scenario 1 (c).

Despite the fact that Figure (5.10) provides insights into the weekly effects of changing promotions, it is difficult to observe whether the actual or the optimal calendar for each product is more profitable. Figure (5.11) shows the sum of the predicted sales volumes for both the actual and optimal promotion calendar. It is observed that the optimal promotion calendar from the retailers' perspective is also optimal for *upc* 2 and *upc* 3. On the contrary, the sales volume of *upc* 1 has declined by 7.000 hectoliters which is 11%. This is explained by the different number of promotions. The total consumption under the optimal promotion calendar is excepted to increase since *upc* 2 gains 14.000 hectoliters (18,6%) and *upc* 3 fortunes a substantial amount of more than 21.000 hectoliters which is a gain of 46%. Manufacturers of *upc* 2 and 3 will exploit this outcome to negotiate the promotion calendar with the retailer since it is profitable from the manufacturer and retailer's perspective, hence the FMCG chain. Contrarily, the result of the optimization problem is not of interest to the manufacturer of *upc* 1 and will investigate strategies that are profitable for the manufacturer and broadens the market in the interest of the retailer.

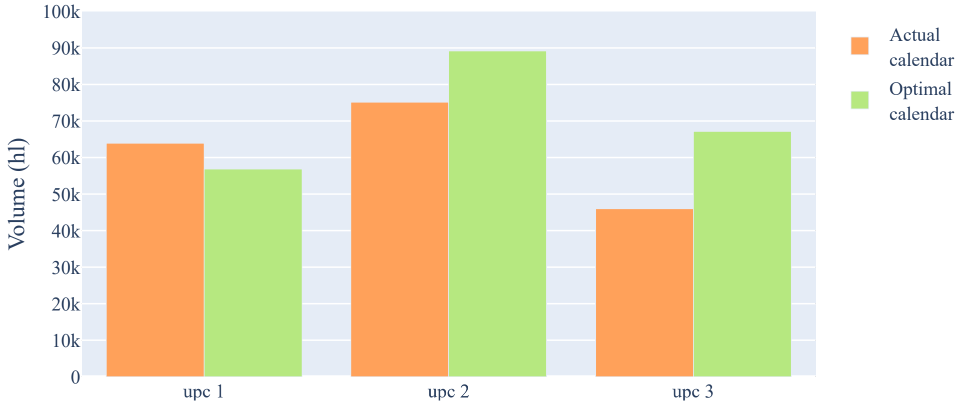


FIGURE 5.11: Total predicted volume of actual and optimal promotion calendar in Scenario 1.

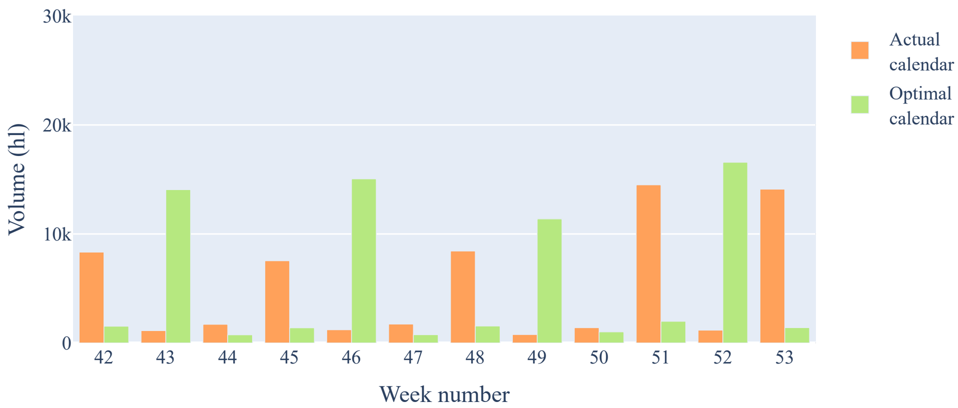
Scenario 2

The optimal promotion calendar of Scenario 2 is shown in Table (5.8). In Table (5.3b), Scenario 2 suggests having 4, 4, and 5 promotional weeks for upc 1, 2, and 3 respectively. This scenario excludes overlapping promotion weeks of two or more products, hence the number of promotions per product is not in accordance with the actual promotion calendar which has to be considered when comparing the predicted sales volumes from the optimal calendar with the actual calendar. The reason for the adjustment is that business constraints take priority over any preliminary data that may contain impurities. Despite the different preliminaries in Scenario 1, the promotion cycle observed is the same which starts with a promotion week of *upc* 3, followed by *upc* 2, and ends with *upc* 1, before promoting *upc* 3 again. Since coinciding promotion weeks are omitted, each product is offered at the base price for two successive weeks between two consecutive promotions.

week no.	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**
***upc* 1**													
***upc* 2**													
***upc* 3**													

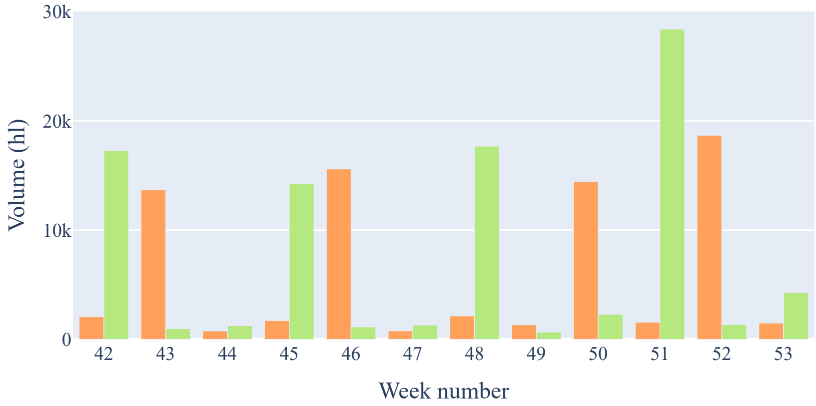
TABLE 5.8: Optimal promotion calendar in Scenario 2 where the red cells are promotion weeks.

Once more, the maximized sales volumes belonging to the optimized promotion calendar are examined. Figure (5.12) presents the predicted weekly sales volume of the optimal and actual promotion calendar, whereas Figure (5.13) shows the sum of the predicted weekly sales volumes per promotion calendar. The figure on weekly sales volume shows that most of the sales volume bars of promotional weeks in the optimal calendar are higher than in the actual calendar. For *upc* 1, all promotional sales are higher in the optimized calendar compared to the actual calendar despite the fact that the actual calendar has one additional promotion week. Nevertheless, the total sales volume of the optimized calendar is above the level of the actual calendar despite one less promotional week. Although the turnover is primarily driven by promotions, it can be concluded that a large number of promotions is not necessarily the best decision for obtaining the largest turnover. Manufacturers tend to avoid choosing a smaller number of promotion slots due to the concern that competitors may opt for additional promotions, which could result in a loss of sales volume and a reinforcement of the cannibalization effect. In this case, *upc* 3 has one promotion week less in the optimized calendar while the total sales volume increases by 21%. The promotion in week 53 contributes significantly since the sales volume in the actual calendar is 3 times the sales of the predicted calendar. A promotion of *upc* 3 alone is more effective compared to sharing a promotion slot with *upc* 1 leading to reduced cannibalization effects on the sales volume of *upc* 3. In the actual as well as the optimized calendar, *upc* 3 is promoted in week 41 which again does not cause a biased comparison between the total volumes in Figure (5.13). Furthermore, the optimal promotion calendar returns a higher predicted sales volume than the actual calendar for all products resulting in an optimal promotion strategy for the entire FMCG chain. In comparison to the promotion calendar of Scenario 1, a calendar without overlapping promotion weeks for different products will not only result in an increase in the total sales volume for the retailer but also for the manufacturers of both *upc* 1 and 2. Although *upc* 3 will gain more in the situation with overlapping promotions, both scenarios result in additional sales.

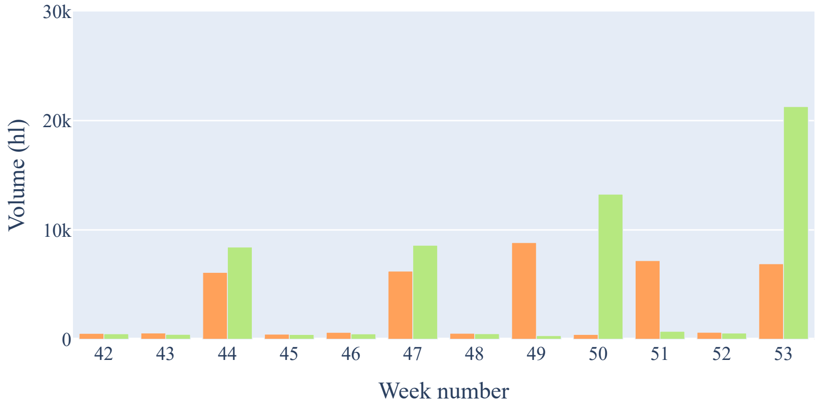


(A) *upc* 1

FIGURE 5.12: Predicted volumes of actual promotion calendar and predicted volumes of optimal promotion calendar in Scenario 2 (a).



(b) *upc* 2



(c) *upc* 3

Figure 5.12: Predicted volumes of actual promotion calendar and predicted volumes of optimal promotion calendar in Scenario 2 (b-c).

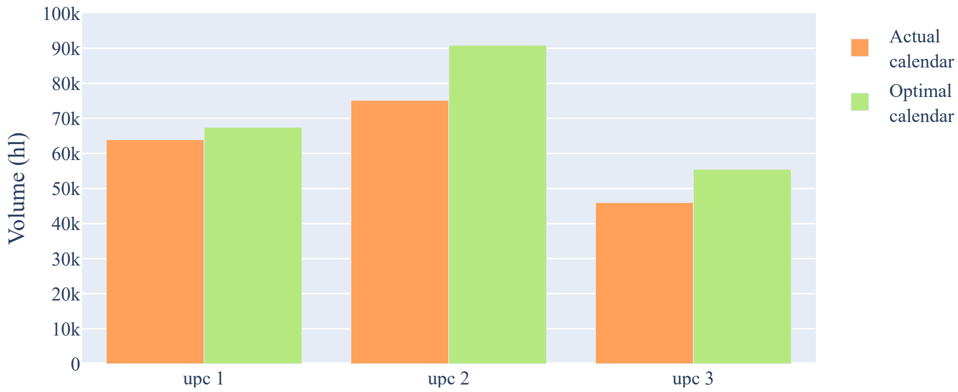


Figure 5.13: Total predicted value of actual promotion calendar and total predicted value of optimal promotion calendar in Scenario 2.

Since the difference in the number of promotions between Scenarios 1 and 2 is one promotion week of *upc* 1 and *upc* 3, two additional instances are created. In Appendix B.2 the preliminary variables and output are presented. Overall, the total sales pattern does not differ from Scenario 1 since the sales volume of *upc* 1 is still below the predicted volume of the actual calendar, and *upc* 2 and 3 gain significantly higher sales, leading to an overall market growth of 15% compared to the actual situation.

Figure (5.14) presents the difference between the total sales volume of the actual calendar and the optimal calendar per instance. Each scenario has a higher turnover than the actual calendar. The FMCG market in Scenario 2 gains 0.28% more compared to Scenario 1 . Although most turnover is driven by promotion pricing, it does not guarantee the highest market turnover. By evaluating the gap in the number of promotions between scenarios 1 and 2, it can be seen that two different reductions in the number of promotions of two products have a diminished impact on the market or an increased impact on the market compared to Scenario 1. Although additional instance 2 would furnish the largest gain in the market, manufacturer 1 will not support this decision. Since manufacturer 1 is the market leader, it is assumed the retailer obeys the manufacturer's opinion by applying Scenario 2.

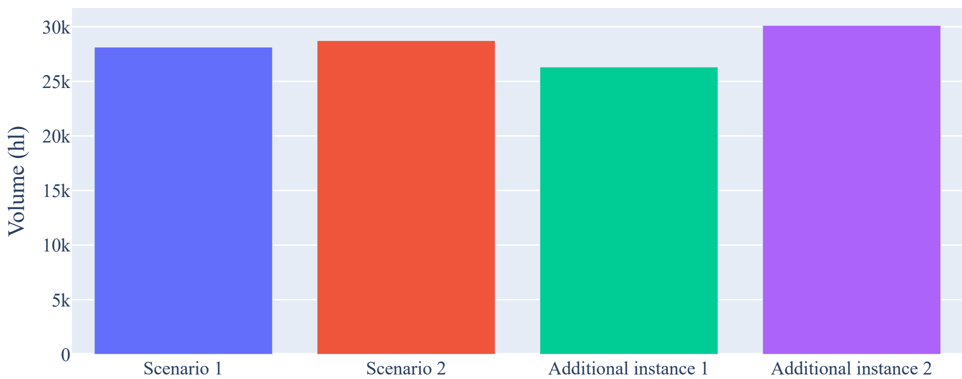


FIGURE 5.14: Difference between the total volume of actual and optimal promotion calendar per scenario.

It is concluded that the results from the regression formula described in Section 5.3.1 are not acceptable for practical use. Hence, it is essential to assess the sensitivity of the optimization problem that employs the estimated parameters of the regression formula. To this end, a 5% adjustment is made to both the negative and positive values making both effects stronger. Although this impacts the predicted sales volumes of the products, the ratio remains unchanged resulting in the same calendar as the optimal promotion calendar. Altering the parameters by varying percentages, yet still within the range of realistic effects, does not result in modifications to the optimal promotion calendar.

The final subject addressed is the complexity of the optimization problem. Although the mathematical problem involves many binary variables, the root relaxation is quickly solved. The branch and bound method is then employed to obtain an integer solution for the problem. The root relaxation is solved in 0.01 seconds for both scenarios. Scenario 1 finds an optimal solution in 47 seconds, taking an auxiliary 18 seconds to prove optimality. For Scenario 2, the branch and bound method was able to find an optimal solution in only 9 seconds, without requiring additional time to verify its optimality. The optimal and best-bound objectives have a 0% gap in both scenarios.

It is crucial to examine the source of the computational costs. Firstly, the conflicting constraints are temporarily removed, that is constraints (4.5g). The root node is solved in 0.05 seconds. However, the branch and bound method requires a lengthy amount of time to find an optimal solution. Applying the promotion schedule from Scenario 1 as input to the optimization problem, the algorithm takes more than 5 minutes to retrieve an optimal solution. Furthermore, there is a considerable gap of approximately 80,000 hectoliters (27%) between the objective value of the root node and the optimal solution found. A larger number of promotions lead to a greater computational time due to a broader feasible solution set. Nevertheless, the conflicting constraints do not lead to complexity or an increase in the computational time since it makes the feasible region smaller and finds an optimal solution faster.

Larger instances involve a greater number of binary variables, resulting in longer computational times in comparison to smaller instances. A problem involving 9 products on a time horizon of 10 contains 90 integer variables. Although the root relaxation is solved in 0.02 seconds, after 180 seconds the branch and bound method has not reached an optimal solution. The optimal solution found has a gap of 220% compared to the best bound, indicating that the MILP is not optimally solved. Furthermore, the difference between the objective value of the root node and the optimal bound is substantial, measuring 7.6 million compared to 1.6 million. To investigate whether the enlargement in the computational cost is solely due to the growth of the number of binary variables, the constraints on piecewise linearization of the exponential function are temporarily ignored such that $\ddot{y}_{it} = \hat{y}_{it}$. Although this does not result in the correct optimal solution, the root relaxation solves in 0.02 to an integer solution. This implies that the complexity arises from the piecewise linear constraints.

Chapter 6

Conclusion, discussion and recommendations

6.1 Conclusion & discussion

In this study, a model for an optimal promotion strategy for FMCG adjusted for cannibalization effects is developed. The regression model in Chapter 3 predicts the sales volume accounting for cannibalization effects as a result of the promotion of products. The optimal promotion calendar is retrieved by solving the optimization problem presented in Chapter 4.

Aggregating the transaction data at the store level is convenient for estimating the cannibalization effects between products and optimizing the promotion calendar with fewer decision variables. However, it is important to note that cannibalization effects may vary across different regions due to differences in brand loyalty and brand preference. This is primarily evident in the beer market, where generally citizens of Amsterdam, Rotterdam, and the surrounding urban area prefer Heineken beers, while residents of Twente favor Grolsch. Furthermore, aggregation leads to a decline in the number of data points, while conversely, a larger number of data points with a similar pattern improves the goodness-of-fit when estimating the coefficients. Not only the amount of data points might result in inaccurate predictions, but also the time horizon is short since the data set contains transactions of one year. The seasonal index variable is determined using the dependent variable of the complete data set, which may lead to a target leak if the calculated seasonal effect depending on the actual sales volume is used as input to fit the regression model and predict the sales volume. This also affects the train-test split, because training the model on 75% of the year and testing the model on the last 25% of the year, leads to a different data pattern between the train and test set. To improve the model accuracy and estimations of the cannibalization effects, one year of transaction data along with external data on seasonal effects such as weather forecasts and festivities can be used to estimate the seasonal effects and train the regression model, while the transaction data of another year can be used to test and validate the regression model.

The forecast accuracy measurements the MAPE and the R^2 are used to analyze the results of the regression model presented in Section 5.3.1. At first sight, the measurements suggest that the model fits the transaction data of beer while returning accurate sales volume predictions. The coefficients have significant p-values, the R^2 is close to 1, and no indications of overfit. Nevertheless, the high MAPE calculated on hectoliters reveals a significant percentage deviation from the actual sales volume due to which the forecast accuracy measure has to be revised resulting in the development of the categorical R^2 on average, the weighted average of data points, and the weighted average of volume. From the results, it follows that the regression model for the non-promo cluster is not functioning based on the negative R^2 for all products while the promo cluster does have a positive R^2 but does not reach the forecast accuracy desired by the manufacturers. The categorical R^2 on average and weighted average of data points are close to zero or negative which indicates that the estimated parameters do not fit the data well, resulting in a regression model that is not applicable to practice since manufacturers aim to achieve a forecast accuracy of 90%. Also, the categorical R^2 on the weighted average of volume does not reach the desired accuracy. Hence, it is concluded that manufacturers rather use current models that do not consider cannibalization effects than a poor-performing prediction method that accounts for cannibalization effects. If the regression model would predict sales volumes following the data pattern of the actual sales volumes driven by promotion prices and cannibalization effects, manufacturers would value the model and use it to acquire an optimal promotion strategy. The regression model has regression variable median sales which does not differentiate between the median sales of promo and non-promo sales. However, including a constant input variable that represents the difference between the median sales in promotional and non-promotional weeks may not lead to a more accurate or distinct prediction, as this effect is already captured in the coefficient of the promotion variable for the product. Therefore, including additional variables that correlate with the sales volume can improve the forecast accuracy. However, correlation with other regression variables should be avoided to prevent the model from multicollinearity, and the number of additional variables should be limited to prevent the regression model from overfitting.

The main goal of this research is to model an optimal promotion calendar for FMCG adjusted for cannibalization effects. Therefore, the regression model is developed and used as input for the promotion calendar optimization problem. Although manufacturers will not use the regression model in practice due to inaccurate sales volume predictions, the model can be used to assess the mathematical program for optimizing the promotion calendar. For the analysis, the predicted sales volumes of the actual calendar are compared with the predicted sales volumes of the optimal calendar resulting from the mathematical program. In Section 5.3.2, the results of the mathematical program are discussed for two scenarios where preliminary input for constraints is retrieved from the data and created by business rules. The results show that the optimal calendars of both scenarios have an increased total predicted sales volume compared to the total predicted sales volume of the actual calendar due to the optimal calendars having the same but another promotion cycle than the actual calendar. When comparing the total sales volume of the two optimal calendars, it is observed that Scenario 2 has an incremented sales of 0.28% compared to Scenario 1 due to a reduced total number of promotions. Although the difference between the total predicted sales volume of Scenario 1 and 2 is only 0.28%, the most important reason to opt for Scenario 2 is that, next to *upc* 2 and 3, also the sales volume of *upc* 1 increases by 5.5%. A constraint can be added to ensure that the total sales volume of a manufacturer is at least the same level in the actual situation to guarantee a Pareto optimal promotion calendar if there exists a feasible solution. To conclude, Scenario 2 is optimal for the FMCG chain for each of the manufacturers compared to the actual situation because it is restricted that competing products share a promotion week. That is why it can be concluded that a reduced number of promotions leads to declined cannibalization effects on the predicted sales volume of the products. If products do not share a promotion week, they do not have to share the consumer's demand which is expected not to change significantly if 1 or 2 products are promoted.

Furthermore, it can be concluded that the optimal promotion calendar is not significantly sensitive to strengthening the impact of the coefficients in the regression model. It is expected that the promotion cycle remained unchanged since it was the same for different scenarios. In fact, an adjustment of the coefficients does not return an altered optimal calendar under identical business constraints. Hence, the influence of the regression model on the optimal promotion calendar is not leading to a major change making it acceptable to utilize the imperfectly fitting regression model. However, if the cannibalization effects estimated in the regression model change drastically, it is expected that the promotion cycle and optimal promotion calendar are revised.

Mathematical programs are complex to solve when many binary variables are involved and it is computationally costly to prove optimality by the branch-and-bound algorithm. The conflicting constraints are not causing the model complexity, in fact, it reduces the feasible region and is computationally efficient. For application in real-world situations, the optimization problem should be solved for more than 3 products leading to an increased number of binary variables. Moreover, the solving time increases exponentially by including more variables which create importance to researching the model complexity. The larger instance that was solved without the linearization constraint suggests that the model's complexity, resulting from linearizing the exponential function, leads to a computationally expensive optimization program. However, one running example does not prove it is generally true. Nevertheless, for a limited number of products or product groups on a time horizon of a quarter, the optimization problem can be solved within an acceptable time constraint for practical use.

Overall, this research contributed a forecast accuracy providing insights into the predictions of clustered data. Moreover, it first attempted to use cannibalization effects in promotion planning, although the regression model could be further improved.

6.2 Future research recommendations

This section discusses several areas of research that can be explored to further improve the optimal promotion strategy for FMCG adjusted for cannibalization effects. In Section 6.1 the aggregation of the transaction data at the store level is discussed implying that cannibalization effects differ per region. Although the promotion planning should be applied nationwide, an accurate regional sales forecast considering regional cannibalization effects can be aggregated to obtain a national promotion plan. However, an incremental of the computational cost of the regression model and optimization problem is expected and nonetheless, more data is essential to achieve functional sales predictions. Moreover, the accuracy of the regression model can be improved by including additional regression variables that are correlated with the sales volume. The price is a factor that is strongly correlated with the sales volume but might be correlated with the promotion variables since it depicts if a product is promoted.

In addition to improving the forecast accuracy of the regression model, there is potential to enhance the insights into the measurements of forecast accuracy. In this study, the data points are clustered based on their promotional status to achieve a promo and non-promo R^2 which are combined in a categorical R^2. However, for each product, there are two or three promo data points located closer to the non-promo cluster which suggests that other clustering methods, such as k-means clustering for a fixed number of clusters or DBSCAN where the algorithm decides on the number of clusters, should be investigated. Furthermore, the determination of promotional transactions is based on the benchmark of the promotion price which is not a robust method. Therefore, sales volumes and price comparisons in the week before and after are criteria that can be added to determine the benchmark of the promotion price.

Finally, the complexity of the optimization problem can be further researched. The running example highlights that the linearization constraint can lead to a computationally expensive optimization problem. Therefore, it is recommended to evaluate the performance of the optimization problem by varying the number of intervals for the linearization of the exponential function and asses the computational cost as well as the deviation of the piecewise linear approximation. In addition, the optimization problem could benefit from the inclusion of additional constraints that limit the feasible region. This approach has the potential to enhance the solution space and improve the model's ability to generate more effective solutions. Moreover, non-linear solving techniques and heuristics can be examined.

References

1. Ali, Ö., Sayın, S., Van Woensel, T., & Fransoo, J. (2009). Sku demand forecasting in the presence of promotions. *Expert Systems with Applications*, *36*(10), 12340–12348.
2. Arsham, H., & Lovric, M. (2011). Bartlett's test. *International encyclopedia of statistical science*, *1*, 87–88.
3. Basson, L., & Kilbourn, J., P.J.and Walters. (2019). Forecast accuracy in demand planning: A fast-moving consumer goods case study. *Journal of Transport and Supply Chain Management*, *13*(1), 1–9.
4. Blattberg, R. C., & Wisniewski, K. J. (1989). Price-induced patterns of competition. *Marketing science*, *8*(4), 291–309.
5. Blattberg, R., & Neslin, S. (1993). Sales promotion models. *Handbooks in Operations Research and Management Science*, *5*, 553–609.
6. Bogomolova, S., Dunn, S., Trinh, G., Taylor, J., & Volpe, R. (2015). Price promotion landscape in the US and UK: Depicting retail practice to inform future research agenda. *Journal of Retailing and Consumer Services*, *25*, 1–11.
7. Chicco, D., & Warrens, G., M.J.and Jurman. (2021). The coefficient of determination R-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*, *7*, e623.
8. Cohen, M., Kalas, J., & Perakis, G. (2021). Promotion optimization for multiple items in supermarkets. *Management Science*, *67*(4), 2340–2364.
9. Cohen, M., Leung, N., Panchamgam, K., Perakis, G., & Smith, A. (2017). The impact of linear optimization on promotion planning. *Operations Research*, *65*(2), 446–468.
10. De Myttenaere, A., Golden, B., Le Grand, B., & Rossi, F. (2016). Mean absolute percentage error for regression models. *Neurocomputing*, *192*, 38–48.
11. Elmaghraby, W., & Keskinocak, P. (2003). Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management science*, *49*(10), 1287–1309.
12. Faria, J. R., & Novak, A. J. (2001). A dynamic model of cannibalism.
13. Farizal, F., Dachyar, M., Taurina, Z., & Qaradhawi, Y. (2021). Disclosing fast moving consumer goods demand forecasting predictor using multi linear regression. *Engineering and Applied Science Research*, *48*(5), 627–636.
14. Fujisawa, T., & Kuh, E. S. (1972). Piecewise-linear theory of nonlinear networks. *SIAM Journal on Applied Mathematics*, *22*(2), 307–328.
15. Greenleaf, E. (1995). The impact of reference price effects on the profitability of price promotions. *Marketing science*, *14*(1), 82–104.

16. Heij, C., de Boer, P., Franses, P., Kloek, T., van Dijk, H., et al. (2004). *Econometric methods with applications in business and economics*. Oxford University Press.
17. Heineken. (2022). Heineken n.v. annual report 2021. https://www.annreports.com/heineken/heineken-ar-2021.pdf
18. Hosseini, F., S.H Zadeh, Shafiee, M., & Hajipour, E. (2020). The effect of price promotions on impulse buying: The mediating role of service innovation in fast moving consumer goods. *International journal of business information systems*, *33*(3), 320–336.
19. Huang, T., Fildes, R., & Soopramanien, D. (2014). The value of competitive information in forecasting fmcg retail product sales and the variable selection problem. *European Journal of Operational Research*, *237*(2), 738–748.
20. Huang, T., Fildes, R., & Soopramanien, D. (2019). Forecasting retailer product sales in the presence of structural change. *European Journal of Operational Research*, *279*(2), 459–470.
21. Ivanov, S. H. (2007). Conceptualizing cannibalization: The case of tourist companies. *Yearbook of International University College*, *3*, 20–36.
22. Katzenelson, J. (1965). An algorithm for solving nonlinear resistor networks. *The Bell System Technical Journal*, *44*(8), 1605–1620.
23. Kuhn, S. (1997). Prisoner's dilemma. https://plato.stanford.edu/Entries/Prisoner-Dilemma/
24. Lopes, R. (2011). Kolmogorov-Smirnov test. *International encyclopedia of statistical science*, *1*, 718–720.
25. Maddah, B., Ben Abdelaziz, F., & Tarhini, H. (2021). Bi-objective optimization of retailer's profit and customer surplus in assortment and pricing planning. *Annals of Operations Research*, *296*, 195–210.
26. Mason, C., & Milne, G. R. (1994). An approach for identifying cannibalization within product line extensions and multi-brand strategies. *Journal of Business Research*, *31*(2-3), 163–170.
27. McColl, R., Macgilchrist, R., & Rafiq, S. (2020). Estimating cannibalizing effects of sales promotions: The impact of price cuts and store type. *Journal of Retailing and Consumer Services*, *53*, 101–982.
28. Mehrdar, A., & Li, T. (2021). Should price cannibalization be avoided or embraced? A multi-method investigation. https://ssrn.com/abstract=3908037
29. Meredith, L., & Maki, D. (2001). Product cannibalization and the role of prices. *Applied Economics*, *33*(14), 1785–1793.
30. Moorthy, K. S., & Png, I. P. L. (1992). Market segmentation, cannibalization, and the timing of product introductions. *Management science*, *38*(3), 345–359.
31. Netessine, S., & Taylor, T. A. (2007). Product line design and production technology. *Marketing Science*, *26*(1), 101–117.
32. Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M., Barrow, D., Taieb, S., Bergmeir, C., Bessa, J., R.J.and Bijak, Boylan, J., et al. (2022). Forecasting: Theory and practice. *International Journal of Forecasting*.
33. Radev, R. (2014). Price planning process in multi-product companies from fast moving customer goods sector. *Economic Alternatives*, *1*, 42–51.
34. Schober, P., Boer, C., & Schwarte, L. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768.

35. Sherali, H., & Adams, W. (2009). A Reformulation-Rinearization Technique (RLT) for semi-infinite and convex programs under mixed 0-1 and general discrete restrictions. *Discrete Applied Mathematics*, *157*(6), 1319–1333.
36. Subramanian, S., & Sherali, H. (2010). A fractional programming approach for retail category price optimization. *Journal of Global Optimization*, *48*, 263–277.
37. Van Heerde, H., Leeflang, P., & Wittink, D. (2002). How promotions work: Scan* pro-based evolutionary model building. *Schmalenbach Business Review*, *54*, 198–220.
38. Van Heerde, H., Leeflang, P., & Wittink, D. (2004). Decomposing the sales promotion bump with store data. *Marketing Science*, *23*(3), 317–334.
39. Villas-Boas, J. (1995). Models of competitive price promotions: Some empirical evidence from the coffee and saltine crackers markets. *Journal of Economics & Management Strategy*, *4*(1), 85–107.
40. Yeoman, I. (2012). Cannibalization. *Journal of Revenue and Pricing Management*, *11*(4), 353–354.

Appendix A

Derivation of OLS parameters

The regression formula for product $i \in P$ is given by

$$\mathbf{y}_i = X_i \mathbf{b}_i + \mathbf{e}_i \tag{A.1}$$

with

$$\begin{aligned} \mathbf{y}_i^{'} &\in \mathbb{R}^T \\ \mathbf{e}_i^{'} &\in \mathbb{R}^T \\ \mathbf{b}_i^{'} &\in \mathbb{R}^k \\ X_i &\in \mathbb{R}^{k \times T} \end{aligned}$$

where K is the number of regressors.

Ordinary least squares (OLS) minimizes the sum squared residual. Rewriting Equation (A.1) to

$$\mathbf{e}_i = \mathbf{y}_i - X_i \mathbf{b}_i$$

such that minimizing

$$\begin{aligned} S_i(\mathbf{b}_i) = \mathbf{e}_i^{'} \mathbf{e}_i &= (\mathbf{y}_i - X_i \mathbf{b}_i)^{'} (\mathbf{y}_i - X_i \mathbf{b}_i) \\ &= \mathbf{y}_i^{'} \mathbf{y}_i - \mathbf{y}_i^{'} X_i \mathbf{b}_i - \mathbf{b}_i^{'} X_i^{'} \mathbf{y}_i + \mathbf{b}_i^{'} X_i^{'} X_i \mathbf{b}_i \end{aligned}$$

gives the estimated parameter values $\mathbf{b}_i$.

Therefore the derivative of $S(\mathbf{b}_i)$ has to be determined and set equal to zero to find the optimum value. The second-order derivative has to be greater than zero in order to check if the optimum is a minimum.

$$\frac{\partial S_i}{\partial \mathbf{b}_i} = -2X_i^{'} \mathbf{y}_i + 2X_i^{'} X_i \mathbf{b}_i$$

First order condition

$$-2X_i^{'} \mathbf{y}_i + 2X_i^{'} X_i \mathbf{b}_i = 0 \implies \mathbf{b}_i = (X_i^{'} X_i)^{-1} X_i^{'} \mathbf{y}_i$$

provided that $(X_i^{'}X_i)^{-1}$ exists. That is, $rank(X_i) = k$. The second-order condition

$$\frac{\partial^2 S_i}{\partial \mathbf{b}_i \partial \mathbf{b}_i^{'}} = 2X_i^{'}X_i$$

which is a positive definite matrix. Hence, $\mathbf{b}_i = (X_i^{'}X_i)^{-1}X_i^{'}\mathbf{y}_i$ is the parameter vector minimizing the sum squared residuals.

Appendix B

Additional Results

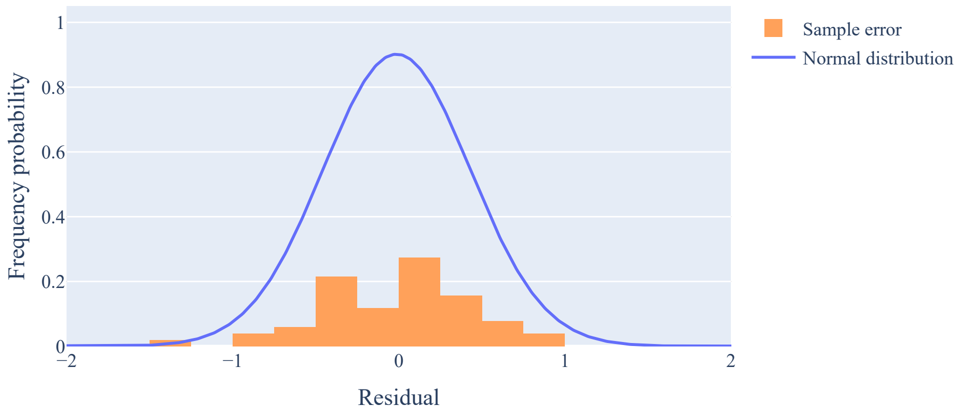
B.1 Regression output and statistics

	Products		
	upc 1	*upc* 2	*upc* 3
seasonal effect	0.5563^{***}	1.0723^{***}	0.8156^{**}
median sales	1.0453^{***}	1.0691^{***}	1.1054^{***}
promo upc 1	1.4789^{***}	-0.5458^{***}	-0.8043^{***}
promo lag upc 1	-0.2744	0.0654	0.2361
promo upc 2	-0.2131	1.5694^{***}	-0.8766^{***}
promo lag upc 2	0.5554^{***}	-0.4476^{***}	-0.0859
promo upc 3	-0.3845^{***}	-0.3405^{***}	2.1907^{***}
promo lag upc 3	-0.1258	-0.0047	-0.1295
N	40	40	40
D	0.0917	0.1676	0.1483
B	1.4735	0.0096	0.0045
d	2.3940	2.3224	2.0820

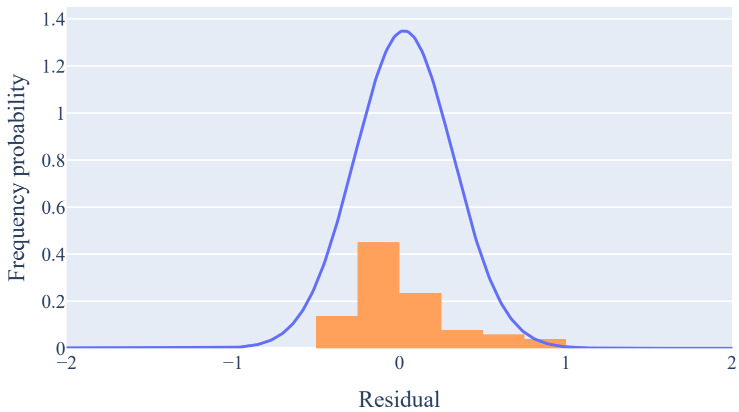
Nota: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

TABLE B.1: Regression coefficients and significance tests.

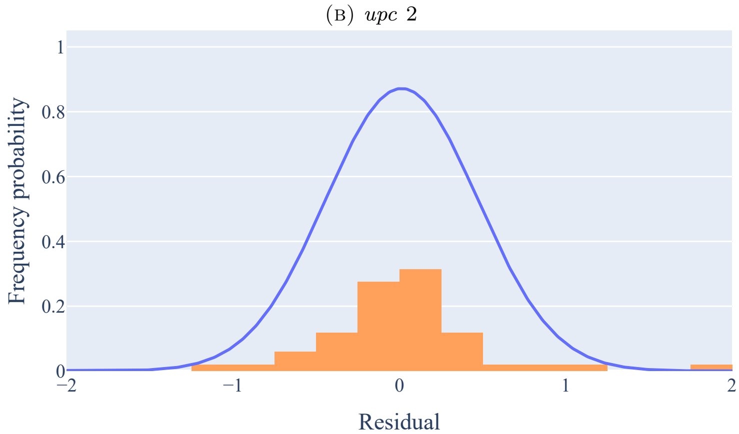
In Table (B.1) the estimated regression coefficients of Equation (3.1a) are shown as well as the significance tests for the regression described in Section 3.2.3. Coefficients with p-values below 0.1 are considered significant, indicating the null hypothesis should not be rejected. The null hypothesis assumes that the independent variable and dependent variable are correlated. The null hypothesis is rejected by the K-S test if the test statistic D is greater than 0.1865. Therefore, it is not rejected that the error distribution is normally distributed for all products. Bartlett's test is used to assess the homoscedasticity of the variables, where the null hypothesis states that the variance of random samples drawn from the error vector is homoskedastic. The test yields $B < 13.8484$, hence the null hypothesis is not rejected. Additionally, since the DW test statistic d is close to 2, the null hypothesis that states no autocorrelation between the error terms is not rejected.



(A) *upc* 1



(B) *upc* 2



(C) *upc* 3

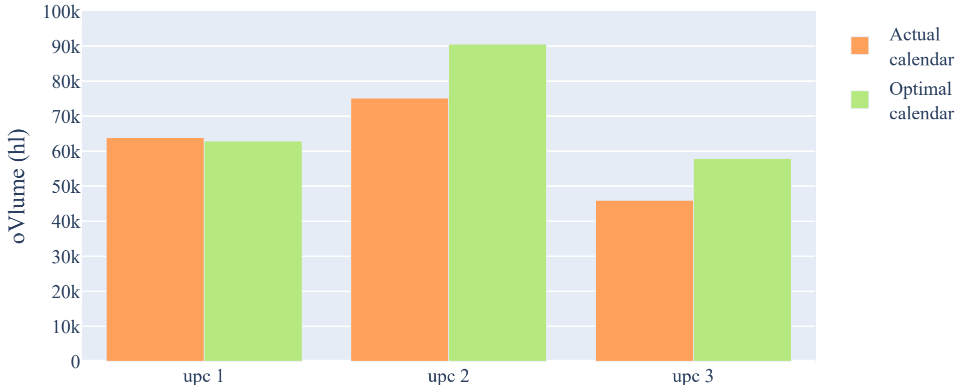
FIGURE B.1: The visualization of the distribution of the error term which supports the outcome of the K-S test with the high likeliness that the error terms are normally distributed with mean μ and standard deviations 0.44, 0.30, and 0.46 for *upc* 1, 2, and 3 respectively.

B.2 Promotion calendar optimization problem: additional instances

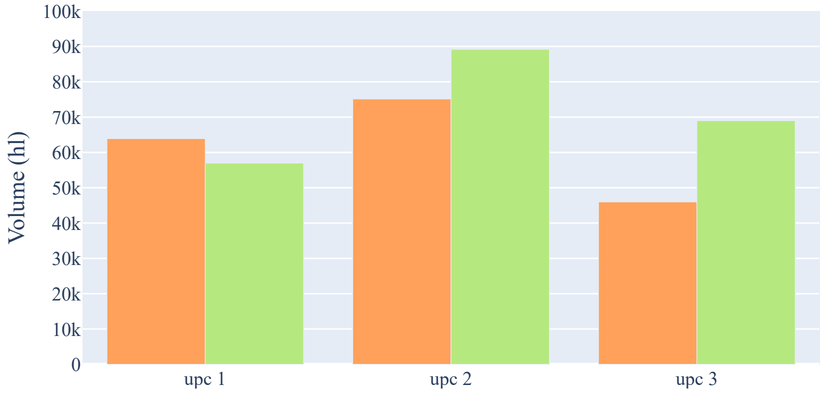
	upc 1	*upc* 2	*upc* 3	Number promotions Additional instance 1	 Additional instance 2
upc 1	0			4	5
upc 2	1	0		4	4
upc 3	0	1	0	6	6

TABLE B.2: Additional instance 1 and 2.

The input of the additional instances 1 and 2 are presented in Table (B.2) and the total predicted sales volume per product for the actual and optimal calendar are shown in Figure (B.2). The optimal calendar of additional instance 1, with one reduced promotion week of *upc* 1 compared to Scenario 1, shows the same promotion cycle as Scenario 1 and 2 having 1 overlapping promotion week of *upc* 1 and 3 in week 43. The total predicted sales volume in Figure (B.2a) shows that the predicted sales volume increased compared to Scenario 1 but is not above the volume of the actual calendar. Figure (B.2b) presents the total predicted sales volume of additional instance 2 where *upc* 3 has one promotional week less compared to Scenario 1. It is observed that the predicted sales volume of *upc* 1 is less compared to additional instance 1 while *upc* 3 gains 20% sales volume and *upc* 2 almost remains the same. Although the optimal calendar has the same promotion cycle as in additional instance 1, the calendar differs because it starts with the promotion of a different product and has an overlapping promotion week in week 47, which affects the cannibalization effects on the products and ultimately leads to changes in their sales volume.



(A) One reduced promotion week of *upc* 1 compared to Scenario 1.



(B) One reduced promotion week of *upc* 3 compared to Scenario 1.

FIGURE B.2: Total predicted sales volume of the actual and optimal promotion calendar, for additional instances 1 and 2 which are an elaboration of scenario 1.