FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

The art of the deal: Machine learning based trade promotion evaluation

David Branco Viana

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Beatriz Brito Oliveira

External Supervisor: Andreia Filipa da Silva Cardoso

July 21, 2021

Abstract

Trade promotions are complex marketing agreements between the retailer and manufacturer, in order to drive up sales. The retailer proposes numerous sales promotions for which the retailer is to help pay for, through discounts and deductions. In the Portuguese consumer packaged goods (CPG) sector, the proportion of price-promoted sales to regular-priced sales has increased to a very significant level such that proper promotional planning is crucial, should the manufacturer's margins withstand the market tendencies.

In this context, a decision support system (DSS) was developed to aid in the promotional planning process of two key product categories of a Portuguese CPG manufacturer. The DSS allows for the planning and simulation of promotional scenarios, providing the manufacturer's commercial team with forecasts of promotional sales, in order to better evaluate a proposed trade promotion and negotiate its terms. The simulation is powered by a forecasting model specific to each retailer-category pair that estimates sales for a given promotion, based only on data available to the manufacturer, which does not have access to scanner-level or even store-level sales data.

Resumo

A fim de oferecer promoções aos consumidores finais, os retalhistas combinam com os produtores descontos e rebates a receber, a fim de cobrir parte dos custos associados. No setor português de bens de consumo, as vendas promocionais são uma porção muito significativa das vendas totais, o que faz do planeamento promocional um processo crucial e determinante para as margens dos produtores.

Neste contexto, um sistema de apoio à decisão (SAD) foi desenvolvido para apoiar o processo de planeamento promocional de duas categorias de produtos de um produtor português de bens de consumo. O SAD permite o planeamento e simulação de cenários promocionais, permitindo à equipa comercial do produtor avaliar um acordo proposto pelo retalhista e negociar melhor os termos de condições deste. A simulação é feita usando um modelo preditivo para cada par retalhista-categoria que estima as vendas provenientes da promoção acordada, baseando-se apenas em dados disponíveis ao produtor, que não inclui dados de vendas ao nível do agregado familiar nem ao nível da loja.

Acknowledgments

I would like to begin by expressing my deepest appreciation for the invaluable guidance given to me by my supervisor, Professor Beatriz Oliveira, whose attentiveness and kindness I am very grateful for.

I am grateful to Andreia Cardoso, Joana Teixeira and Paulo Sousa for having worked alongside me on this project, taking me under their wing and mentoring me. I also thank the rest of my colleagues at LTPlabs, for welcoming me with open arms into an exceptional workplace.

To my amazing parents, my lovely sister and my family, I thank you for your unending love and unconditional support. These last five years were filled with ups and downs, through all of which you've constantly cheered me on.

To my classmates and close friends, I am grateful for your comradery and I wish to be with you all again very soon, hopefully closer than 6 feet apart.

David Viana

"A danger foreseen is half-avoided."

Cheyenne Proverb

Contents

List of Figures

List of Tables

List of Algorithms

Abbreviations and Symbols

CPG	Consumer Packaged Goods
DSS	Decision Support System
EAN	European Article Number
EDLP	Everyday Low Price
ERP	Enterprise Resource Planning
GBM	Gradient Boosting Machine
HDPE	High-density polyethylene
HILO	High-Low
MAD	Mean Absolute Deviation
MAPE	Mean Absolute Percentage Error
MSE	Mean Square Error
MSRP	Manufacturer's Suggested Retail Price
PET	Polyethylene terephthalate
PRT	Promotional Registry Tool
PSPS	Promotional Scenario Planner and Simulator
RMSE	Root Mean Squared Error
SKU	Stock Keeping Unit
SR	Scenario Runner
USD	United States Dollar
VBA	Visual Basic for Applications
WBIAS	Weighted Bias
WMAPE	Weighted Mean Absolute Percentage Error
WTP	Willingness To Pay

Chapter 1

Introduction

Over the last half decade, both the intensity of price discounts given as part of trade promotions between retailers and manufacturers and the share of total sales they represent have been rapidly growing. The four largest British supermarkets reported that, in 2016, in the consumer packaged goods (CPG) sector, 45% of the average consumer's expenditure was on price-promoted products (Hill, 2016). Two major Australian supermarkets reported that, on average, in any given week of 2017, they price-promoted 26% to 30% of all products in the beverage category (Zorbas et al., 2019). Nielsen found that promotional sales accounted for nearly half of the Portuguese CPG sector's sales in 2018, approximately three times as much as for the Spanish counterpart's sales (Nielsen, 2020). With so many trade promotions being proposed by the retailer to the manufacturer, it becomes paramount to assess how profitable or worthwhile a given promotion will be.

1.1 Purpose

In this context, a CPG manufacturer is interested in having a custom Decision Support System (DSS) developed, to aid in the promotional planning process for its main product categories, namely the olive oil and vegetable oil categories. The main objective of this work is to successfully develop such a DSS as well as train models that can accurately predict promotional sales for the various retailers and categories, despite only leveraging the limited data available to the manufacturer. This work was developed in a consulting firm, in a project with a Portuguese CPG manufacturer.

1.2 Document structure

This dissertation is divided in several chapters. Chapter 2 provides a literature review, both to fill the reader in on key concepts that are used throughout the rest of the document and to present the state-of-the-art for promotional sales forecasting. Chapter 3 describes the company and the problem it faces, as well as the retailers it works with and the market it operates in. Chapter 4 details the decision support system implemented, as well as the data used and the models trained on such data. Chapter 5 elaborates on the models obtained, their performance and metrics as well as comparing these results, drawing conclusions about the different retailers and categories and describing the final deployment of the system. Chapter 6 concludes the document, summarizing the purpose, methodology and contributions of this work, as well as pointing to some possibilities of future work.

Chapter 2

Literature Review

This chapter aims to fill the reader in on key concepts that are used throughout the present document. It starts with some important economics concepts that permeate the remainder literature review, the central one being the price elasticity of demand. Afterwards, the area of Marketing is covered, where some basic concepts are presented, as well as vital marketing terms, related to manufacturers and their relationship with both retailers and consumers, heavily used throughout the work. Subsequently, the field of study of forecasting is explored, emphasizing its techniques and methods, where time series methods and machine learning models are included. Modern forecast methods require a solid data framework to support them, such that the final section covers the field of data mining, an indispensable supporting branch of knowledge, with regards to treating the data as well as managing the database that holds it.

2.1 Economics

As explained in a later section, manufacturers and retailers make deals among themselves to offer lower prices to customers, to influence demand upward. Demand, or the demand curve, is defined as the amount customers can and want to purchase of some good or service at each price point, with quantity demanded referring to a specific point in this curve (Greenlaw and Shapiro, 2017). The law of demand states that there is inverse proportionality between demand and price, all other variables being equal. This is quite an intuitive idea, since, when offered a lower price, a given consumer might become able or even tempted to purchase more, increasing quantity demanded. There are exceptions to this law, namely, Veblen goods and the more controversial Giffen goods. Veblen goods include luxury items, whose higher price translates into exclusivity and subsequential desirability, low cost items that sell less, since they transmit an idea of inferior quality to the consumer, and items such as stocks, where rising stocks and falling stocks are perceived as more and less desirable, respectively, by the average investor. Giffen goods are necessity items that, when priced higher, are more highly demanded in extremely poor communities. This happens because increasing the price of a basic good, a cheap source of calories such as bread, for example, means the poor individual cannot afford to complement it with something more expensive, such as meat, and will have to resort to consuming more bread for survival (Jensen and Miller, 2007).

Willingness to pay (WTP) is the maximum price a customer is willing to pay for a good or service (Stobierski, 2020). WTP is a dynamic characteristic of a customer, influenced by various phenomena such as supply shortages or the emergence of stronger competitors, for example (Stobierski, 2020).

The price elasticity of demand is the relative change in demand produced by a relative change in price (Greenlaw and Shapiro, 2017).

$$\varepsilon(p) = \frac{\Delta_{\%} d}{\Delta_{\%} p} \tag{2.1}$$

A given good is elastic if $|\varepsilon(p)| > 1$ and inelastic otherwise. If $|\varepsilon(p)| = \infty$, the good is perfectly elastic and, if $|\varepsilon(p)| = 0$, the good is perfectly inelastic, that is, variations in price do not affect demand. Both retailers and manufacturers are interested in knowing how reducing price will increase demand, therefore estimating the price elasticity of demand is desirable in this context.

2.2 Marketing

This work deals with several Marketing concepts about sales and manufacturer-retailer dynamics, which involve both consumer and trade promotions, both of central importance to the work.

2.2.1 Baseline sales, incremental sales and sales lift

Baseline sales, also called normal sales, are defined by Blattberg et al. (1996) as "the estimate of sales after controlling for and/or removing the effects of specific marketing activities". Incremental sales are the sales generated through marketing programs (Farris et al., 2010) and add to the baseline sales to form the total amount of sales. Sales may be expressed in the number of units sold or in gross revenue from the sale of those units. Sales lift is the increase in sales, relative to the baseline, normally expressed in percentage (Farris et al., 2010).

$$\text{Sales Lift} = \frac{\text{Incremental Sales}}{\text{Baseline Sales}} \tag{2.2}$$

The higher the sales lift, the more incremental sales were induced by a given marketing promotion, and the more successful, that is, profitable the latter can be considered (Farris et al., 2010). The profit generated by the promotion can then be calculated as the number of incremental units sold times the promotional unit profit minus the marketing costs associated with the promotion.

2.2.2 The manufacturer-retailer dynamics

Sales promotions are marketing programs that incentivize retailers or consumers to purchase the product being promoted. On the consumer side, sales promotions, also called consumer promotions, involve sample offering, couponing, premiums such as a "buy one get one free", vouchers, immediate price reductions and other incentives (Broderick and Pickton, 2005; Abraham and Lodish, 1987).

When directed at trade customers, such as retailers, sales promotions are also called trade promotions (Broderick and Pickton, 2005). Trade promotions can take the form of allowances (such as discounts, additional free of charge products, special terms), point-of-sale materials or joint promotions, among other enticing terms (Broderick and Pickton, 2005; Abraham and Lodish, 1987; Blattberg and Levin, 1987). Joint promotions are a partnership between the retailer and the manufacturer, involving sales promotions of any kind, where the costs and efforts are borne by both parties (Broderick and Pickton, 2005). Joint promotions generally involve a contractual agreement, whose retailer-side benefits are conditional on the retailer providing a minimum amount of value to the manufacturer, by offering customers a price reduction on certain items or by boosting advertisement efforts on the manufacturer's products (Abraham and Lodish, 1987). Manufacturers may encounter difficulties ensuring that the retailer honors certain parts of the deal, especially with regards to advertisement efforts, whereas sales promotions are far easier to monitor (Abraham and Lodish, 1987) and collect data upon.

When an allowance is given to a retailer, as a standalone trade promotion or as part of a joint promotion, the degree to which the discount offered to the retailer is passed onto the customers is called the pass-through degree.

$$\text{Pass-through} = \frac{\text{Customers' total savings due to the trade promotion}}{\text{Retailer's total savings due to the trade promotion}} \tag{2.3}$$

Opposite to passing the discount through, the retailer may also take advantage of the allowance and stockpile, also known as forward buying, which is an important source of profit for the retailer (Blattberg and Neslin, 1989) but does not directly benefit the manufacturer nor the customers.

It is important to distinguish two types of sales, from the manufacturer's point of view. Firstly, sell-in is defined as the number of units which are sold to retailers by the manufacturer for resale to consumers. Secondly, sell-out, also called sell-through, is defined as the number of units which are sold by retailers to consumers (Vitasek, 2006).

For simplicity, henceforth the immediate price reduction sales promotions will be referred to as sales promotions, and joint promotions will be referred to as trade promotions, unless written otherwise. In this context, promotional intensity is defined as the depth of the immediate discounts offered by the retailers and promotional frequency as the percentage of time these discounts are offered.

2.3 Forecasting

Bil Keane, an American cartoonist famous for his newspaper comic *The Family Circus*, is also famous for the following quote:

> *"Yesterday is history, tomorrow is a mystery, today is a gift of God, which is why we call it the present."*

Undeniably, future events are mysterious, of both random and predictable nature. It is this predictable nature that forecasting aims to harness into foresight. This section covers time series methods, for dealing with historical data, and causal methods, that, besides historical data, leverage other variables that may provide useful insight into the future, in which machine learning and its forecasting methods are included, which will be extensively used in this work.

2.3.1 Definition and types of forecasting methods

Forecasting uses historical data and useful knowledge of any future events to make accurate predictions about the future, called forecasts (Hyndman and Athanasopoulos, 2018). Two distinct types of forecasting methods exist, namely, qualitative and quantitative, the latter being applicable only if past numerical data is available and there is valid reason to believe there is a pattern or mechanism underlying such data. Qualitative methods, also called judgmental methods, are based on opinions, generally being used when quantitative methods cannot be applied, as can be the case, for example, in launching new products. Examples include the Delphi method, that aggregates the anonymous opinion of various experts, or forecasting by analogy, where the situation which is object of analysis is compared to an analogous one, in hopes that the driving factors are the same and therefore a forecast by proxy can be made, an example being real estate appraisal by comparison (Hyndman and Athanasopoulos, 2018). Regarding quantitative methods, two main types exist, namely, time series methods and causal methods.

2.3.2 Time series methods

A time series is defined as data collected sequentially over time (Box et al., 2015). Time series analysis attempts to detect patterns in the data and to generate forecasts using past values (Cryer, 1986).

Stationary models are a class of stochastic models that assume the mechanism behind the process behavior does not change over time (Cryer, 1986). It can then be said that a stochastic process is stationary if the mean and variance are constant over time (Box et al., 2015).

Nonstationary models, on the other hand, are better fit for modeling most business and industry phenomena such as weekly sales or monthly units produced (Box et al., 2015). Nonstationary time series include trends (long-term increase or decrease in the time series), seasonal patterns (rises and falls of a known fixed frequency, due to seasonal factors, such as the time of the year or the day of the week) or cycles (rises and falls of a irregular frequency, usually due to economic conditions) in the data (Hyndman and Athanasopoulos, 2018). Often nonstationary time series have to be transformed into stationary ones to allow the use of stationary models, differencing the data, that is, computing the differences between consecutive observations, being one such transformation (Hyndman and Athanasopoulos, 2018).

Some simple time series forecasting methods

The average method calculates the future value to be equal to the average of past values:

$$\hat{y}_{T+h|T} = \bar{y} = \frac{1}{T}\sum_{k=1}^{T} y_k \tag{2.4}$$

where $y_1, \ldots, y_T$ is the historical data and $\hat{y}_{T+h|T}$ is the h-step forecast, taking into account all observations up to time T (Hyndman and Athanasopoulos, 2018).

The naive method sets the future value equal to the last observation. As simple as it is, it is the optimal forecast when the data behavior is that of a random walk (Hyndman and Athanasopoulos, 2018).

$$\hat{y}_{T+h|T} = y_T \tag{2.5}$$

The seasonal naive method sets the future value equal to the last observation of the analogous season:

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)} \tag{2.6}$$

where m is the seasonal period and $k = \lfloor (h-1)/m \rfloor$.

Simple exponential smoothing lays somewhere in-between the naive method and the average method, by using the smoothing parameter α to emphasize the more recent values:

$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1} \tag{2.7}$$

$$\hat{y}_{t+h|t} = \ell_t \tag{2.8}$$

$$\tag{2.9}$$

2.3.3 Causal methods

Causal methods use the historical values of the variable to be forecast, alongside predictor variables to perform their analysis (Hyndman and Athanasopoulos, 2018). These variables, also called features, are factors that have, or are believed to have, a causal relationship with the time series, in the same manner as marketing programs or competitor initiatives affect future sales (Chambers et al., 1971).

2.3.4 Machine learning

It is in this context of predictor variables that machine learning stands out. Machine learning algorithms analyze data and attempt to learn how they are related to each other, with learning problems spanning multiple categories. A supervised learning model is one where the training data includes both the input vectors and their corresponding target vectors (Bishop, 2006). If there are a finite number of categories to match each input vector, the model is performing classification, otherwise, if one or more continuous variables are associated with each input vector, the model is performing regression (Bishop, 2006). Besides supervised learning, the basic machine learning paradigms include unsupervised learning, where the algorithms strive to structure the data, and reinforcement learning, where the algorithm can perform actions in an environment and adjust its behavior based on the feedback they receive.

In a regression problem, a number of features correspond to a numeric variable, called the target, whose relationship to the features is to be approximated via a regression model. Generally, the dataset, containing the aforementioned features and target, is split in three partitions, namely, the training, validation and test dataset. The training dataset is used to fit the model, iteratively, in a way that minimizes a function that indicates how well the model, given the corresponding features, estimates the target. This function is called the loss function, $L(t, y(\mathbf{x}))$, which can be any function, but generally takes the form of $C(y(\mathbf{x}) - t)^2$, with $C > 0$, the square assuring the loss is always non-negative and that negative and positive losses do not average out (Bishop, 2006).

A sufficiently complex model can memorize the training set, such that the loss nears 0. Not because the model has not learned the underlying relationship between the features and the target but because the model has become overfit to the training set, which means the model will have poor foresight capability. To prevent this, the models are tested on the validation dataset, since they validate that the model has not become overfit by evaluating the model's performance in a dataset it has not yet seen. This validation performance is then used to select the best model among the ones trained. Since the model validated is biased towards the validation dataset, the test dataset performance is used as the real-world benchmark of the model (Bishop, 2006).

Many machine learning techniques for regression exist, of which two of interest to this work are listed below. Both are based on decision trees, models which split the feature space into regions, via successive, recursive binary splits, each of which is governed by a simple model, normally a constant (Bishop, 2006). The root node begins with the whole dataset, and selects the feature that best splits it, according to a given metric, dividing the space into two regions. These two regions are then recursively split, until a stopping criteria has been met.

Random forests A learning method developed by Breiman (2001) for both classification and regression problems. The method uses an ensemble of decision trees, which together form a strong model. The uniqueness of the method comes from the conjunction of bootstrap aggregation, also called bagging, and random feature selection. Bagging samples the original training set with replacement, generating different datasets to increase the diversity of the trees trained. Random feature selection restricts each split to only use a random subset of features, further reducing the correlation between trees.

Gradient boosting machines A learning method developed by Friedman (2001), gradient boosting machines (GBM) was initially called "multiple additive regression trees". GBM successively trains decision trees on the residuals left by the previous trees, correcting their deficiencies, which is shown by Friedman (2001) to be a combination of both gradient descent and boosting. Gradient descent is an iterative algorithm for finding the local minimum of a differentiable function, by successively calculating the gradient at any point in said function and stepping in the opposite direction to it (Cauchy, 1847). Boosting is a meta-algorithm that combines the outputs of many weak learners into one strong learner.

2.3.5 Evaluating a model's performance

In training a model, the differences between the predicted values and the target values are called residuals, also called errors in accuracy measure methods. When testing the model, the difference between the predicted value and the actual value is called the forecast error (Hyndman and Athanasopoulos, 2018).

Various accuracy methods are used to evaluate and compare different models. Scale-dependent errors like the mean absolute deviation (MAD), the mean square error (MSE) or the root mean squared error (RMSE) are not suitable for comparing models with different units (Hyndman and Athanasopoulos, 2018). Percentage errors are unitless and include the most commonly used mean absolute percentage error (MAPE). Problems arise with MAPE if any y_n[1] is zero, since it leads to division by zero errors, which may inhibit its use in certain cases. To measure a model's bias, BIAS, also called mean percentage error (MPE), is used.

An additional metric, R^2, also called the coefficient of determination, assesses the goodness of fit of a given model. It accomplishes this by comparing the squared residuals of its predictions for a given dataset with the squared residuals of produced by a model that predicts the average of said dataset (Devore, 2008). The latter model would have an R^2 value of 0, while a worse model would have a negative R^2 value and the ideal model would have an R^2 value of 1. The formulas for all

[1]Here t is switched to n, to denote a generalization of the expressions to values that do not follow a sequential, time-series related logic.

the above metrics are described below:

$$\mathrm{MAD} = \frac{1}{N}\sum_{n} |\hat{y}_n - y_n| \tag{2.10}$$

$$\mathrm{MSE} = \frac{1}{N}\sum_{n} (\hat{y}_n - y_n)^2 \tag{2.11}$$

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} \tag{2.12}$$

$$\mathrm{MAPE}\ (\%) = \frac{100}{N}\sum_{n} \left|\frac{\hat{y}_n - y_n}{y_n}\right| \tag{2.13}$$

$$\mathrm{BIAS}\ (\%) = \frac{100}{N}\sum_{n} \frac{\hat{y}_n - y_n}{y_n} \tag{2.14}$$

$$R^2 = 1 - \frac{\sum_n (y_n - \hat{y}_n)^2}{\sum_n (y_n - \bar{y})^2} \tag{2.15}$$

To differentiate between errors on low-selling items and high-selling items, weighted metrics are used, namely, a custom variant to MAPE, WMAPE (weighted MAPE), and to BIAS, WBIAS. Instead of using the actual value as the weight, a generic weight W_n is used, chosen according to the type of target being predicted, as shown in Eq. 2.16 and Eq. 2.17.

$$\mathrm{WMAPE}\ (\%) = 100 \cdot \left[\sum_{n} \frac{|\hat{y}_n - y_n| \cdot W_n}{y_n}\right] / \left[\sum_{n} W_n\right] \tag{2.16}$$

$$\mathrm{WBIAS}\ (\%) = 100 \cdot \left[\sum_{n} \frac{\hat{y}_n - y_n \cdot W_n}{y_n}\right] / \left[\sum_{n} W_n\right] \tag{2.17}$$

2.4 Data mining

Data mining is a data analytic discipline that focuses on the discovery of structures and patterns in large, complex datasets. Modern data mining combines statistics, machine learning, database technology and other data analysis technologies to find similarities or anomalies in the dataset, known as pattern detection, as well as to summarize the dataset or to predict future outcomes, both part of what is known as model building (Hand and Adams, 2014).

Data mining is a complex and iterative process, as shown in Fig. 2.1. First, a dataset has to be selected from the database(s), followed by cleaning, where inconsistencies and outliers are removed, and the data is manipulated to assure consistency among itself, especially if multiple databases are involved (Bose and Mahapatra, 2001). Afterwards, the dataset undergoes preprocessing, also called feature extraction, where the original input variables are transformed into more suitable representations for ensuing data analysis (Bishop, 2006). This data analysis phase is where one or more models are trained, in order to generate interesting patterns, from which insights and knowledge can then be extracted after careful interpretation and evaluation.

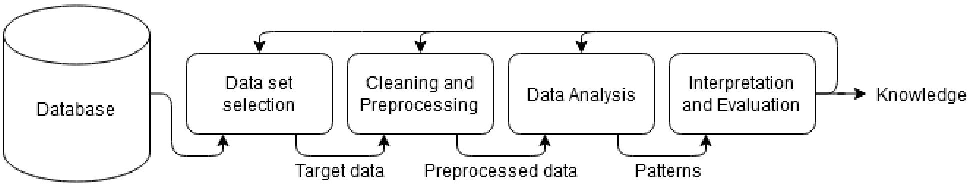

Figure 2.1: Overview of the data mining process. Adapted from Bose and Mahapatra (2001).

2.5 Sales promotion evaluation

As far as the authors can tell, no literature covers the topic of sales promotion evaluation without the use of scanner-level data. However, ample research has been done on the retail side, with the use of scanner-level data and other retailer specific data, where two articles stand out. Cooper et al. (1999) implemented a sales promotion forecasting system, in order to aid a large retailer to plan promotions effectively. The system leverages daily or weekly scanner-level data from multiple stores and consumer panel data, as well as promotional features such as the type of ad used and the product display allocated. Forecasting was done through linear and log-linear models. Divakar et al. (2005) developed a large forecasting model inserted in a DSS for a billion-dollar revenue CPG company with extensive use of scanner-level data among other data sources. Analogously to the DSS implemented by Cooper et al. (1999), sales forecasting was done through linear and log-linear models, using a wide array of features, including display sizes for each brand, and prices of multiple products across brands. Abolghasemi et al. (2020) expanded on both of these articles, testing various models, including various time-series and machine learning models. Unlike its predecessors, less data was available such that only historical sales and price data was used. They ultimately show that the volatility of demand has a significant impact on forecasting accuracy and that simple statistical models can outperform more complex models when this volatility is present.

Chapter 3

Problem Description

A consumer packaged goods manufacturer, faced with increasing promotional frequency and intensity, is concerned about how effective its promotional planning is, considering how significant promotional sales are in the total sales figure. With a substantial number of sales promotions held on a monthly basis, the manufacturer finds it difficult to assess each promotion's impact. Higher levels of promotional frequency translate into increased promotional saturation, reducing the novelty factor of a sale. Consequently, the average impact of a given promotion diminishes, making it harder for the average promotion to be worthwhile. Considering the promotional saturation, this work aims to assess where it is possible to detect bad deals before they happen, using the manufacturer's insider knowledge and available data.

3.1 The company

The manufacturer considered in this work deals with numerous retailers, in several countries, offering multiple product lines across various CPG categories. This work focuses on the Portuguese market and on two CPG categories, olive oil and vegetable oil, where the manufacturer is the market leader, more so on the latter. Regarding vegetable oil, the company supplies roughly 75% of the total market demand in liters, 40% of which is through their flagship brand, the remainder split among store brands[1] across multiple retailers. Although the olive oil market is more heavily disputed, the manufacturer controls around 45% of the total market demand, split roughly evenly between their brands and store brands. Vegetable oil and seeds are sourced from multiple farmers, all over the world, whereas the olive oil is extracted from olives produced in thousands of acres' worth of olive groves owned by the company. The raw materials are treated, transformed into oil, and packaged, ready to be sold, mostly to retailers, wholesalers and the hospitality industry, also known as HORECA (the Dutch and French abbreviation for Hotel/Restaurant/Café). Additionally, the company recently began to explore selling direct to customer, by opening its first retail store. Product advertising to the masses is done mostly through retailers: via sales brochures, TV spots

[1] Store brand products, also called own brand products, are produced by a manufacturer for resale under a brand controlled by a retailer.

and radio commercials. Sales promotion activities include simple immediate discounts, loyalty-card-only discounts, quantity based discounts (such as "buy one get one free"), coupons, among others. For confidentiality reasons, the manufacturer's identity will be concealed, being referred simply as "the manufacturer" or as "the company".

3.2 The retailers

The manufacturer's olive oil and vegetable oil lines are present in all relevant retailers, which buy the products either directly from the manufacturer or via third-party cash and carry wholesalers. However, not all sources of revenue are equally significant nor do they lend themselves as easily to the analysis this work intends to do. Cash and carry wholesalers are responsible for a sizable portion of the company's brand product sales, however, data is scarcely available. Furthermore, the manufacturer provides store brands for many retailer, for both categories, which represents a significant source of revenue outside the scope of this work, since no promotional planning is done by the company for this type of product.

Some retailers practice a "Everyday Low Price" (EDLP) strategy while others practice a "High-Low" pricing strategy (HILO) where temporary discounts are offered on occasion. Retailers with the former strategy are the focus of this work, since EDLP retailers rarely engage in joint promotions, while HILO retailers heavily rely on them to both maintain customer loyalty and compete with other retailers.

The work focuses on two key retailers, whose identities are concealed for confidentiality reasons. Each of these two retailers control approximately 20% of the Portuguese market share, with retailer A having provided the company with sell-out data, while for retailer B only sell-in data is available. Retailers A and B make up most of the available promotional plan data, having the most recorded sales promotions for both categories, retailer B contributing with more records than retailer A. The work does not cover the rest of the retailers for two main reasons. Firstly, some of the retailers engage in negligible amounts of trade promotions with the manufacturer. Thus, very few of their sales promotions involve the company's say and investment, such that the benefit of effectively planning such promotions is limited. Secondly, some purchase only a small portion straight from the manufacturer, the rest of which is supplied by third-party wholesalers, which drastically decreases the significance of the company's sell-in data for those retailers, considerably hindering any further analysis.

3.3 The Portuguese market

The Portuguese consumer packaged goods market is dominated by sales promotions where promotional sales accounted for half of the sector's sales in 2018 (Nielsen, 2020). However, manufacturer data indicates their categories are even more price promoted. Insider data shows that the company's promotional frequency has at least doubled in the last 5 years[2], as shown in Table 3.1,

[2] All 2020 figures are calculated on data truncated at the end of June 2020.

via the weighted[3] average of actively promoted weeks. Increasingly more sales are discounted, as seen in Table 3.2, and, furthermore, the company's promotional intensity has rapidly risen, tripling or quadrupling over the years for certain retailer-category combinations, as seen in Table 3.3.

	2016	2017	2018	2019	2020*
Olive Oil	13.3%	32.7%	64.6%	78.0%	79.7%
Vegetable Oil	36.0%	33.5%	49.0%	73.4%	78.7%

Table 3.1: Weighted average of actively promoted weeks over the years (using company branded product sales data, in Retailer A).

	Olive Oil	Veg. Oil	Cookies	Cereal	Beer	Yogurts
2016	32.5%	70.1%	-	-	-	-
2017	69.0%	72.6%	-	-	-	-
2018	87.2%	72.8%	43.0%	47.0%	77.0%	54.0%
2019	94.7%	76.0%	46.0%	50.0%	78.0%	68.0%
2020*	91.5%	93.3%	49.0%	50.0%	80.0%	66.0%

Table 3.2: Weighted average of percent promotional sales. The olive oil and vegetable oil figures are based on company branded product sales data, in retailer A, while the other categories' figures are based on sales data of comparable products and retailers.

		2016	2017	2018	2019	2020*
Olive Oil	Retailer A	3.8%	11.4%	26.7%	33.2%	40.6%
	Retailer B	11.3%	15.9%	26.6%	38.0%	44.7%
Vegetable Oil	Retailer A	6.4%	6.7%	6.9%	13.4%	16.4%
	Retailer B	8.5%	10.0%	19.4%	25.7%	26.8%

Table 3.3: Weighted promotional intensity over the years, using company branded product sales data, for both olive oil and vegetable oil.

[3] The weight used was the sales of each product.

3.4 The company's promotional planning process

Management defines the annual volume and margin targets as well as setting the Manufacturer's Suggested Retail Price (MSRP) for the various products in each product line at the end of every year, as part of their annual corporate goal setting process. This target definition takes into consideration current and forecasted commodity prices for the year, for both olive oil and the various kinds of vegetable oil (sunflower, sesame, etc.). On a quarterly basis, the promotional planning team drafts a rough proposal for each one of the retailers, taking into consideration both the sales promotions held in the same quarter of the previous year and quarter management's goals. The draft includes aspects of the sales promotion, such as what combinations of products and discounts will be part of a given promotion, how long it will last, among other factors. The draft also includes trade promotion aspects, such as pass-through, deductions[4], and other financial costs. All of these aspects are subject to heavy weekly negotiation. The proposal is merely a suggestion, with the retailer having the final say on all aspects. However, it is not worthwhile for the retailer to offer a sales promotion without a solid trade promotion behind it, thus the reason why it seeks a compromise with the retailer. The commercial team negotiates based on sales forecast, as well as forecasts of the commodity prices underlying the products, as well as their own expertise. To pressure the company into agreeing to a trade promotion, the retailer often uses upcoming competitor promotions as leverage. After the sales promotion has ended, the team analyze the results using the latest sales data and their experience.

[4]Retailers deduct from the manufacturer's invoice a compensation for their promotional or advertising efforts that the retailer believes to be just compensation.

Chapter 4

Proposed Solution

The solution developed is a system comprised of five elements, namely, the Promotional Registry Tool, the database, the Promotional Scenario Planner and Simulator, the Scenario Runner and the predictive models, as shown in Fig. 4.1. This chapter describes each element in detail, with special emphasis on the predictive models, the focus of this work.

4.1 Decision Support System overview

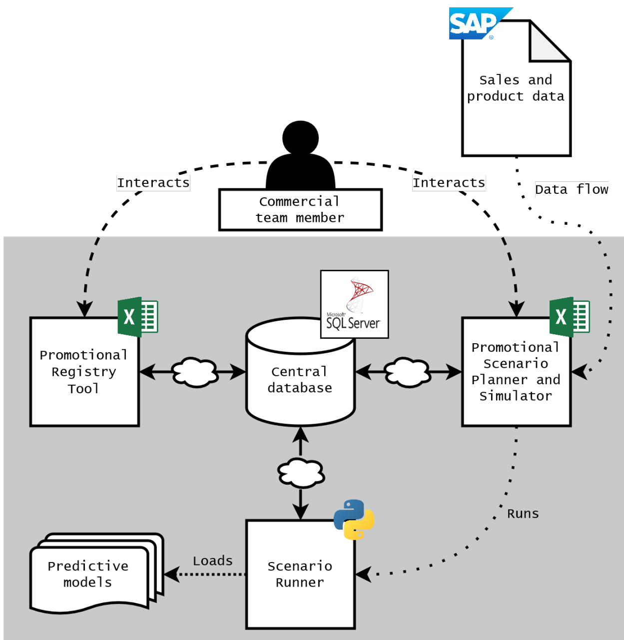

Figure 4.1: Decision support system diagram.

The Promotional Scenario Planner and Simulator (PSPS) is the focal interface of the DSS, the main purpose of which is to simulate future promotional plans, also called promotional scenarios, via the Scenario Runner (SR), which loads and runs a predictive model that estimates the sales of a proposed promotion. The Promotional Registry Tool (PRT) allows the commercial team to register past promotions in a more systematic and robust way, for future access by the PSPS and the SR. The central database, administrated by the company, serves as the connecting block between the PRT, the PSPS and the SR, as well as saving the relevant manufacturer data in an online and structured manner.

4.2 The Promotion Registration Tool

The PRT was developed to aid promotional plan data collection, making it a less error-prone process and streamlining future data flows, two needs identified while treating and importing existing data. The tool was developed in Excel using VBA (Visual Basic for Applications), for ease of development and deployment and due to the commercial team's familiarity with Excel.

Characterizing a promotion

Promotions can be split in two groups, namely, company promotions, which promote the company's products, and competitor promotions, which promote the competitors' products. For all promotions, the following information is captured:

- Start and end date
- Owner (whether it is a company promotion or a competitor promotion)
- Retailer
- Category
- Scope (which products are being promoted and by which fashion they are grouped)
- MSRP
- Discount percentage
- Discount modality (either a direct, loyalty-card-only or "tax-free" discount)
- Promotional price
- Type (the marketing campaign associated with the promotion)
- Geographic coverage (whether it is applied nationwide or to a restricted area[1])
- Special display type (whether the products are displayed in a way that boosts visibility)

A promotion's scope encompasses the manner by which products are grouped together in a promotion, which can range from a single product, capacity-wide, or segment-wide promotion to a brand-wide or category-wide promotion. For company promotions, more scope details are captured, such as combinations of scope (for example, category-capacity-wide) to further narrow down the product selection and, if applicable, the attributes that define the scope such as, for example, capacity or segment. Table 4.1 shows a graphical example of the information registered.

[1] The restricted area could refer to mainland Portugal only, the archipelagos of Madeira and Azores, or a small group of stores.

Owner	Category	Retailer	Geog. Coverage	Type	Scope	Special display type	Promotional price	MSRP	Discount	Start date	End date
Company	Olive Oil	Retailer B	Nationwide	Pamphlet	Capacity-wide (1L)	None	7.50	10.00	25%	7/6/2020	14/6/2020
Competitor A	Olive Oil	Retailer A	Nationwide	Pamphlet	Capacity-wide	None	6.50	10.00	35%	5/6/2020	12/6/2020
Competitor B	Vegetable Oil	Retailer B	Restricted	Weekend	Category-wide	None	3.49	6.98	50%	13/6/2020	16/6/2020

Table 4.1: Example rows of promotional plan data.

4.3 The Promotional Scenario Planner and Simulator

The PSPS is the main deliverable, combining planning, simulation and monitorization capabilities, as well as enabling the commercial team member to import necessary data. From the tool's home screen (see Fig. 4.2), the user can define the retailer, category and the planning horizon with which he intends to work with, in terms of planning and simulation, as well as navigate to the various other screens, which will be detailed next.

Figure 4.2: Home screen.

4.3.1 Planning

The user can plan future promotional scenarios for a given retailer-category pair and a specific planning horizon. The planning process starts with defining template promotions, which are fully defined promotions except for their start and end date, in the screen shown in Fig. 4.3. These template promotions can be scheduled multiple times in week-long slots for a given promotional scenario (see the scenario editing screen shown in Fig. 4.4), simplifying the planning process since most promotions are often repeated over time. The slots do not constrain the promotions' length to be a multiple of a week, nor force them to start on a Tuesday[2], existing only for ease

[2] All slots begin on a Tuesday, the day of the week when most week-long promotions start, in the various retailers.

of visualization. Furthermore, the user can see the historical competitor promotional intensity, a feature explained in detail in section 4.5.2, as well as load a simple, read-only rendition of the promotional plan that took place in the previous year, for the corresponding planning period.

The scenario overview screen lists all scenarios saved, as well as displaying details about each scenario, that is, its name, date of creation, whether it has been evaluated yet, its planning horizon, number of template promotions used and number of promotions scheduled, as shown in Fig. 4.5. Additionally, this screen allows the user to delete scenarios.

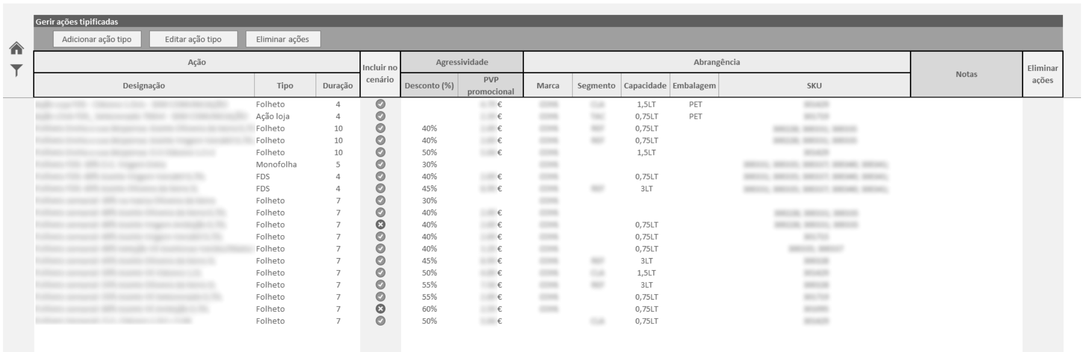

Figure 4.3: Template promotion editing screen.

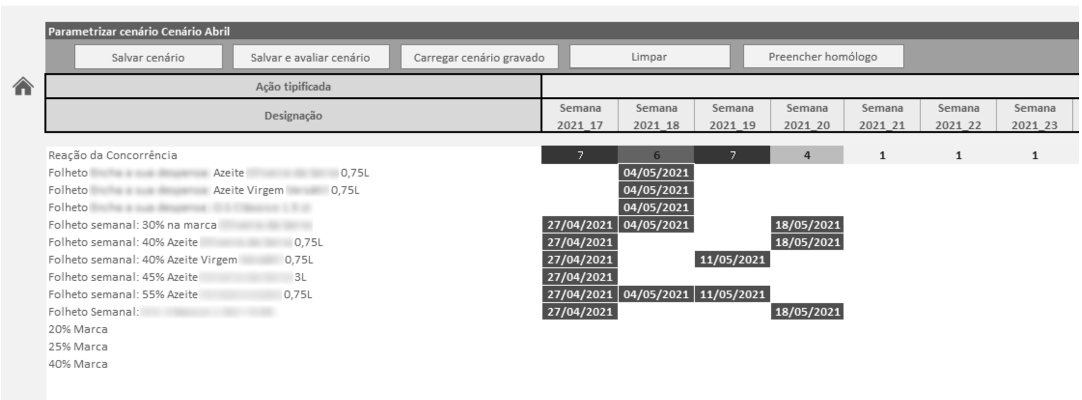

Figure 4.4: Scenario editing screen.

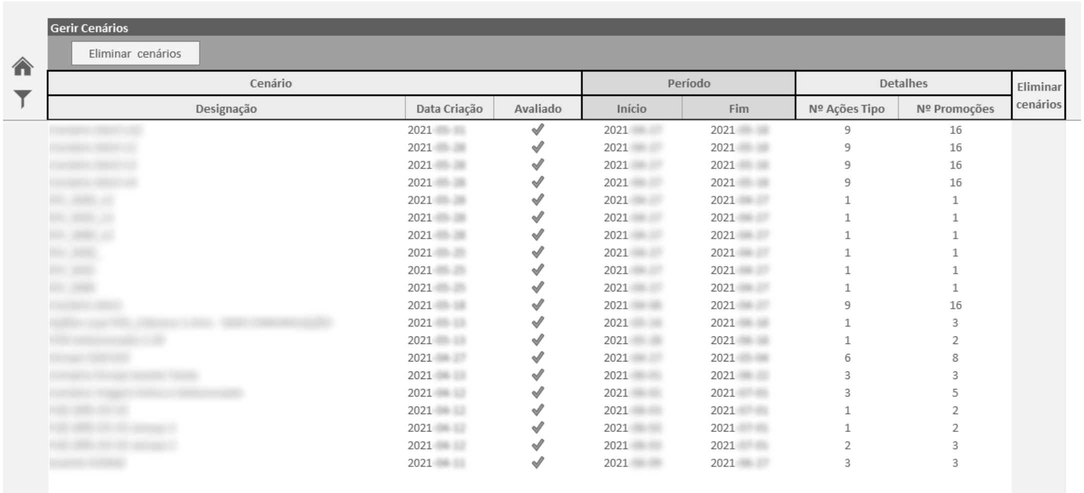

Figure 4.5: Scenario overview screen.

4.3.2 Simulation

The scenario simulation capabilities of the PSPS are powered by the Scenario Runner (SR), a Python executable that runs previously trained predictive models, that is triggered by a button present in the scenario editing screen. A simplified version of SR's process is shown in Fig. 4.6. The SR begins by loading historical data and the scenario to be evaluated, preparing both before feeding them to the model. This preparation involves calculating some of the necessary features, as well as data preparation background work, to ensure appropriate input is fed to the predictive model. It then loads and runs the model, after which the predictions are saved to the database, and control is returned back to the PSPS. Once a scenario has been evaluated, its overall predicted performance can be analyzed, with a performance breakdown of the multiple promotions that comprise it, as can be seen in Fig. 4.7. The PSPS also allows the user to compare multiple scenario evaluations at once.

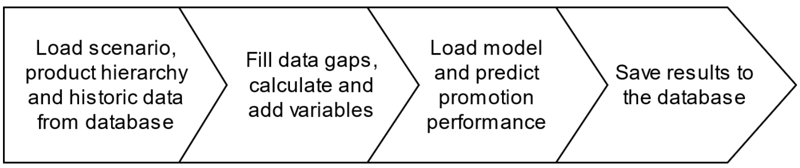

Figure 4.6: Scenario runner process diagram.

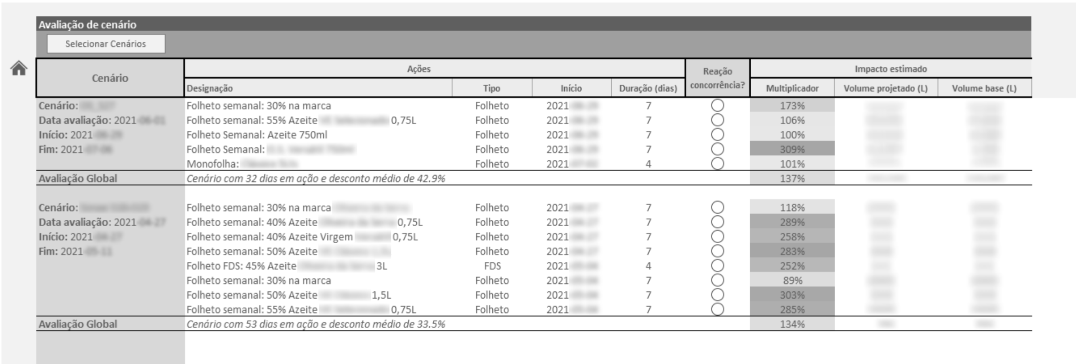

Figure 4.7: Simulation results screen, showing multiple runs.

4.3.3 Monitorization

In keeping with the motto of learning from past experience, the PSPS enables the commercial team members to analyze past promotion performances within the tool, as shown in Fig. 4.8. Unlike the planning and simulation screens, monitorization is not limited to the current retailer-category pair, allowing the user to see all past promotions and their details, as well as filter said promotions based on start and end date, category, retailer, promotion type, discount percentage range and geographic coverage.

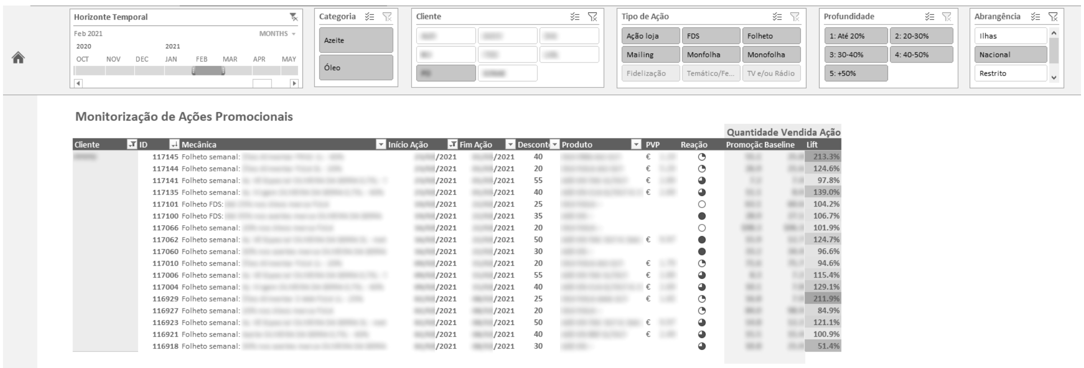

Figure 4.8: Monitorization screen.

4.3.4 Importing and editing data

The PSPS allows for manually importing data to the central database, namely, the sell-in and product hierarchy data, as well as update the MSRP tables and set the retailer stocking periods. The manufacturer uses SAP ERP (Enterprise Resource Planning) software, developed by SAP SE, to export sell-in and product hierarchy data so it can later be imported manually into the PSPS, among other uses. Initially, an automatic SAP ERP data flow to the database was to be implemented, but such was not possible in the short term.

4.4 Data availability and use

The manufacturer made available four key sources of information, namely, sales data, product hierarchy data, promotional plan data and MSRP tables. In terms of sales data, retailer A provided daily sell-out data, spanning from July 2016 to March 2021, while for all retailers daily sell-in data spanned from January 2011 to late May 2021.

Regarding sell-in, correlation is assumed between the liters that arrived to the retailer during the stocking period and the sales resulting in liters from a given promotion. However, this approximation does not hold particularly well for low rotation items, for which the retailer is incentivized to stock an initial amount, often at a reduced price, that sells over a long period of time. Reversely, higher rotation products have a fairly predictable and strong demand, such that the retailer is incentivized to purchase only the necessary amount it has forecasted, since stocking more than a few week's worth of these items is undesirable and costly. Assuming the retailer restocks frequently and just in time for promotions, in order to reduce stockholding costs, correlation studies between the start of a promotion and the total sell-in sales during relative time frames (e.g., 1-3 days, 4-7 days before the promotion) were carried out at the promotion level, in order to determine a stocking period, but no correlation was found. Turning to expert advice, following talks with the commercial team, the stocking period was then defined as the week before the promotion and is adjustable in the PSPS, as mentioned in section 4.3.4. If the retailer stockpiles heavily during the stocking period or before, peaks and troughs of perceived promotional sales occur, respectively, translating into extreme promotional multipliers, which will be covered in detail in section 4.5.1. This was rectified through filtering out promotions whose promotional multipliers were below a minimum of 0.5 and above a maximum of 10, values which were indicated to us by the commercial team to be extraordinary low and high, respectively.

The remainder of this section briefly describes how the various data sources were treated and processed, before being imported to the database.

Sales data Sell-in data includes units returned to the manufacturer, such that some rows had negative sales, which were adjusted to zero sales. Sell-out data required no processing.

Product hierarchy data The data was parsed and inserted into the central database, with the final format containing the following information about each product:

- SKU
- EAN
- Brand
- Category (either vegetable or olive oil)
- Subcategory (e.g., organic olive oil, sunflower seed oil)
- Capacity (in liters)

- Packaging type (PET[3] or HDPE[4] for vegetable oil, either PET or glass for olive oil, the latter denoting a more premium packaging)

To uniquely identify a product in across different countries and manufacturers, GS1, a global non-profit organization, is responsible for issuing Universal Product Codes (UPCs) and European Article Numbers (EANs), among other standard product numbering schemes, for each product to be sold. The EAN is often found in barcode form, which enables retailers to scan the items at checkout or for warehousing needs. The Stock Keeping Unit (SKU) is a code that uniquely identifies a product, as well as its characteristics, inside a company. The same product with different attributes, such as thematic packaging, has an SKU for each combination, so the company can track its different variations, but a single EAN, for consistent external use by the retailer or other entities. This SKU to EAN mapping was done manually and validated by the company's commercial team.

Promotional plan data The various record fields were entered manually, some in a sufficiently structured manner, while others underwent heavy treatment, requiring considerable manual effort to render them machine-readable and usable. Table 4.2 shows a simplified version of the structure of the initial plan data, where each row describes a past sales promotion for a single retailer, for both company promotions and competitor promotions.

Owner	Retailer	Geogr. Coverage	Type	Products involved	Promotional price	Description	Start date	End date
Company	Retailer A	Nationwide	Pamphlet	"Extra Virgin Olive Oil"	3.00	"Pamphlet: Ol. Oil Extra Virgin Olive Fair 750 ml +55% (6.99)"	4 Jan 2020	11 Jan 2020
Competitor A	Retailer B	Archipelagos	Weekend	"Sunflower seed oil"	(null)	"Weekend: All Sun. Oil Praise the Sun Wknd 35%"	12 Jan 2020	19 Jan 2020

Table 4.2: Simplified example rows of rough promotional plan data.

It includes the start and end dates, the type of marketing campaign behind the promotion, the geographic coverage, the promotional price (if the campaign targets a single EAN), as well as two human descriptions, one for the products involved and another for the promotion.

The description of the products involved could detail the brand, subcategory, capacity or any combination of the three. The description of the promotion generally started with specifying the marketing campaign, optionally followed by the name of the promotion and ended with a summary of the promotions' characteristics, as seen in the following example:

> *Pamphlet: "fill up your pantry": 35% discount on a selection of <brand> olive oil (gourmet, select)*

Here, the discount percentage of 35%, the subcategories (Gourmet, Select) and the brand are automatically detected using regular expressions[5] and processed via a small computer script. For company promotions, these details enable the creation of a list of potential SKUs, using the product

[3]Polyethylene terephthalate, a common, multipurpose plastic.
[4]High-density polyethylene, a plastic known for its high strength-to-density ratio.
[5]A regular expression specifies a search pattern with which text is matched against.

hierarchy, which are then crossed with the SKUs active at the time of the promotion, that is, the ones with nonzero total sales during the 6 months prior to the promotion. Using the previously established SKU to EAN mapping, this results in a list of EANs believed to have been included in the promotion, thus the promotion-level row can thereby be exploded into multiple promotion-EAN-level rows.

MSRP tables The MSRP tables provided were vastly incomplete, such that the promotional plan was used to fill-in those gaps, by leveraging promotional price and discount data, where available, to infer the MSRP in practice at the time of that promotion.

A note on the 2020 pandemic

Unusual patterns of purchase were detected in both March and April of 2020, most likely attributed to panic buying, such that these months were not considered for this work. A supposed "new normal" may have altered these general purchase patterns, however, since the categories' products can be classified as essential goods with stable demand, there is little reason to consider the remainder of 2020 and 2021 as having significantly different patterns from the pre-pandemic era ones, era which ended at the start of March 2020, for the purposes of this work.

4.5 Predictive model

Unless explicitly mentioned otherwise, "model" refers to a group of models generated in similar manner for different retailer-category pairs, where only the input data is changed. The crux of the work was training a predictive model to power the simulation capabilities of the PSPS. To do so, the model needed to balance both predictive performance and sensitivity to some key features, since the user is likely to often want to experiment with the system in atypical ways, for example, with discounts rarely observed, historically. A predictive performance driven model, hereinafter called model α, was trained, using a modified algorithm, adapted from the feature selection algorithm described by Kuhn and Johnson (2019), altered to include hyperparameter tuning. The model obtained lacked the sensitivity sought for, leading to the development of a second model, in iterative fashion, which became the model currently in use in the PSPS by the commercial team, at the time of writing, hereinafter referred to as model β. This section describes the various steps taken to arrive at the final model, namely, target definition, feature engineering and the building of both model α and β.

4.5.1 Target definition

Defining the target for the model was not straightforward, as there were many possibilities to explore, given the peculiarity of the high promotional frequency present in both categories and both retailers. This intensity hinders a traditional analysis, which would involve calculating a sales lift on top of a baseline estimate, since with more promotions, less non-promoted data points are available for such estimate (Blattberg et al., 1996). The proposed method abandons the concept of sales lift and instead focuses on establishing the performance of the promotion as a multiple of its average sales, called the promotional multiplier.

Average daily sales and the promotional multiplier

The average daily sales, $\bar{L}_{e,d}$, serves as a pseudo-baseline for each EAN, defined as the rolling average of one year of sales for each day d in liters, lagged by one month:

$$\bar{L}_{e,d} = \frac{\sum_{d' \in D_d} L_{e,d'}}{d_{\max} - d_{\min} + 1}, \quad D_d = \{d - 13 \text{ months}, \cdots, d - 1 \text{ month}\} \tag{4.1}$$

where $d_{\max}$ and $d_{\min}$ are the latest and earliest nonzero sales day included in the averaging period D_d, respectively. This rolling average is done on one full year of data, to avoid seasonality effects, and is used to calculate PM_p, the promotional multiplier for each promotion:

$$\mathrm{PM}_p = \frac{L_p}{(\mathrm{ED}_p - \mathrm{SD}_p + 1) \cdot \sum_{e \in \mathcal{E}_p} \bar{L}_{e,\mathrm{SD}_p}} \tag{4.2}$$

where SD_p and ED_p are the start date and end date of promotion p, respectively, and L_p is the total sales of promotion p, calculated differently for sell-in and sell-out data:

$$L_p = \sum_{d \in D_p} \sum_{e \in \mathcal{E}_p} L_{e,d} \tag{4.3}$$

$$D_p = \{\mathrm{SD}_p, \cdots, \mathrm{ED}_p\} \qquad \text{(Sell-out version)} \tag{4.4}$$

$$D_p = \{\mathrm{SD}_p - \mathrm{STK}_{\text{start}}, \cdots, \mathrm{SD}_p - \mathrm{STK}_{\text{end}}\} \qquad \text{(Sell-in version)} \tag{4.5}$$

The method for calculating L_p using sell-out data is intuitive, since in this case $L_{e,d}$ represents the sales of EAN e on day d to the customers. In the sell-in case, $L_{e,d}$ represents the amount of liters of EAN e delivered to the retailer on day d, from which customers will buy at an uncertain date. According to the sell-in approximation assumption described in section 4.4, the amount of liters sold during the promotion will be roughly proportional to the amount of liters delivered during the stocking period of the promotion, with $\mathrm{STK}_{\text{start}}$ and $\mathrm{STK}_{\text{end}}$ expressing the number of days between the start and end, respectively, of the stocking period and the start date of the promotion.

Target chosen

With the goal of estimating L_p, the model's target is then the average daily sales of EAN e during promotion p in liters, $L_{\text{avg}_{e,p}}$, where e is being promoted in p, therefore generating EAN-level predictions. To arrive at promotion-level predictions, which is the ultimate goal, the various EAN-level predictions are aggregated and transformed according to Eq. 4.6 into an estimate of the promotional multiplier of promotion p, $\hat{\mathrm{PM}}_p$, as defined in Eq. 4.2.

$$\hat{L}_p = (\mathrm{ED}_p - \mathrm{SD}_p + 1) \cdot \sum_{e \in \mathcal{E}_p} \hat{L}_{\mathrm{avg}_{e,p}} \tag{4.6}$$

The model targets the sales of a given EAN e included in a promotion p rather than the sales of the promotion as a whole, in order to be able to capture individual EAN effects. Additionally, estimating the average daily sales reduces the complexity of the model's task, removing the issue of dealing with how different promotion lengths would affect the outcome.

4.5.2 Feature engineering

In an attempt to maximally capture the various aspects of a promotion, substantial effort was put into feature engineering, such that many features, some more significant than others, were created.

Foundational features

The more complex features engineered were built on top of basic features regarding promotion-specific features, namely, the promotion's duration, discount offered, D_p, and the month of its start date, as well as EAN-specific features, specifically, the EAN's brand, segment, MSRP_e and capacity C_e. Additionally, some straightforward features were calculated, namely, the promotional price (Eq. 4.7) and the promotional price per liter of an EAN (Eq. 4.8).

$$\mathrm{PP}_{e,p} = \mathrm{MSRP}_e \cdot (1 - D_p) \tag{4.7}$$

$$\mathrm{PPL}_{e,p} = \mathrm{PP}_{e,p} / C_e \tag{4.8}$$

Company promotion features

The following features focus on relating both concurrent and past company promotions to a given promotion p.

Weight of EAN e on promotion p and weight of promotion p on the category These features allow capturing effects related to the weight in sales of an EAN on the promotion (Eq. 4.9) and the weight of a promotion on its category (Eq. 4.10), for instance, distinguishing best-sellers from niche products and high impact promotions from low impact promotions, respectively. They are also part of many company promotion features, which are covered next.

$$\mathrm{WEP}_{e,p} = \frac{\bar{L}_{e,p}}{\sum_{e' \in \mathcal{E}_p} \bar{L}_{e',p}} \tag{4.9}$$

$$\mathrm{WPC}_p = \frac{\sum_{e \in \mathcal{E}_p} \bar{L}_{e,p}}{\sum_{e \in \mathcal{C}_p} \bar{L}_{e,p}} \tag{4.10}$$

Cannibalization This feature attempts to capture the effect that other EANs being concurrently promoted have on a given EAN e, during promotion p. It does so via a weighted sum of discounts of every EAN-promotion-day combination, for EANs different from the one being analyzed and considering only the days included in promotion p and the promotions that took place on those days, including promotion p, as described in Algorithm 1.

Algorithm 1 Calculating the cannibalization effect faced by EAN e during promotion p.

```
procedure CN_{e,p}
    CN_{e,p} ← 0
    for day d in the active period of p do
        CN_{e,p,d} ← 0
        for EAN e' in all EANs except e do
            CN^{max}_{e,p,d,e'} ← 0
            for promotion p' in all promotions do
                if p' is active on day d and e' is included in p' then
                    CN_{e,p,d,e'} ← D_{e',p'} · WEP_{e',p'} · WPC_{p'}        ▷ See Eq. 4.9, Eq. 4.10
                    CN^{max}_{e,p,d,e'} ← max(CN_{e,p,d,e'} , CN^{max}_{e,p,d,e'})
            CN_{e,p,d} += CN^{max}_{e,p,d,e'}
        CN_{e,p} += CN_{e,p,d}
    return CN_{e,p}
```

Promotional intensity of an EAN This feature allows the model to take into consideration past promotions of a given EAN, via a weighted sum of past discounts, for each day in the month previous to the current promotion. It is calculated in similar fashion to *Cannibalization*, the main differences being that it looks to past promotions rather than concurrent ones and considers only the same EAN, as illustrated in Algorithm 2.

Algorithm 2 Calculating the promotional intensity of EAN e during promotion p.

```
procedure PI_{e,p}
    PI_{e,p} ← 0
    for day d in the month previous to p do
        PI^{max}_{e,p,d} ← 0
        for promotion p' in all promotions do
            if p' was active on day d then
                PI_{e,p,d} ← D_{e,p'} · WEP_{e,p'} · WPC_{p'}        ▷ See Eq. 4.9, Eq. 4.10
                PI^{max}_{e,p,d} ← max(PI_{e,p,d} , PI^{max}_{e,p,d})
        PI_{e,p} += PI^{max}_{e,p,d}
    return PI_{e,p}
```

Promotional intensity of the category Similar to the *Promotional intensity of an EAN*, this feature attempts to capture the overall promotional intensity of the category and is depicted in Algorithm 3.

Algorithm 3 Calculating the promotional intensity of the category during promotion p.

```
procedure PIC_p
    PIC_p ← 0
    for day d in the month previous to p do
        PIC_{p,d} ← 0
        for EAN e in all EANs of the category do
            PIC^{max}_{p,d,e} ← 0
            for promotion p' in all promotions do
                if p' was active on day d then
                    PIC_{p,d,e} ← D_{e,p'} · WEP_{e,p'} · WPC_{p'}        ▷ See Eq. 4.9, Eq. 4.10
                    PIC^{max}_{p,d,e} ← max(PIC_{p,d,e} , PIC^{max}_{p,d,e})
            PIC_{p,d} += PIC^{max}_{p,d,e}
        PIC_p += PIC_{p,d}
    return PIC_p
```

Other features Four simple features were also created, namely, *Days elapsed since last promotion*, *Number of recent promotions* (with recent promotions referring to promotions taking place in the 15 days prior to the promotion in hand), *Number of concurrent promotions* and the *Percentage of EANs included in promotion p* (in relation to the whole category).

Competitor promotion features

As mentioned in section 4.4, some limited competitor promotional plan data was made available by the manufacturer. Unlike with company products, no competitor product hierarchy was available, such that the analysis had to be simpler. In order to capitalize on the available competitor promotional plan data, some features were created in order to capture these effects.

Competitor promotional intensity The competitor promotional intensity of promotion p is determined based on the set of competitor promotions that took place during promotion p's time frame, using, namely, the number of said promotions, their owners, their (maximum) discounts and their promotional scopes. The algorithm for this feature was created using expert knowledge and is as described in Algorithm 4.

Other features Analogous to the competitor promotional intensity, the *Recent competitor promotional intensity* is calculated similarly to the *Competitor promotional intensity*, the difference being that the competitor promotions being analyzed took place in the 15 days prior to promotion p. Additionally, two other simple features were created, namely, the *Days elapsed since last competitor promotion* and the *Number of recent competitor promotions*.

External features The commercial team relayed onto us that discount and promotional policy was dependent on the current price of the commodity underlying the category, as well as forecasts of said commodity, such that two external features were included, namely, the *Commodity monthly*

Algorithm 4 Calculating the competitor promotional intensity during promotion p.

```
procedure CPI_p
    if a category- or brand-wide major competitor promotion is in effect or five major competitor promotions are in effect then
        if the maximum discount of such promotion(s) is greater than 25 then
            return 7
        else
            return 6
    if a segment- or capacity-wide major competitor promotion is in effect or two major competitor promotions are in effect or a category- or brand-wide minor competitor promotion is in effect then
        if the maximum discount of such promotion(s) is greater than 25 then
            return 5
        else
            return 4
    if a major competitor promotion is in effect or a segment- or capacity-wide minor competitor promotion is in effect or five minor competitor promotions are in effect then
        return 3
    if a minor competitor promotion is in effect then
        return 2
    return 1
```

price, in USD[6] per metric ton, and the *Commodity monthly percentage price change*. The commodities chosen were the global price for olive oil and the global price of sunflower oil, for the olive oil and the vegetable oil categories, respectively. The commodities' ticker symbols, that is, their unique identifier in the stock market, are POLVOILUSDM and PSUNOUSDM, respectively, and their data was sourced from the Federal Reserve Bank of St. Louis[7].

4.5.3 Model α

A feature selection algorithm introduced by Kuhn and Johnson (2019) automatically selects features while accounting for the bias introduced by ranking the feature set on the training set. Model α was generated via a modification of this algorithm that includes hyperparameter tuning in the same bias avoiding spirit, as detailed in Algorithm 5. The learning method chosen was the Gradient Boosting Machine (GBM) method, and the hyperparameters to be tuned were the maximum depth, *max_depth*, of each individual tree and the number of trees, *ntrees*, of said GBM model. Alongside GBM, Random Forests and other readily available algorithms, such as linear regressions and deep learning algorithms, were experimented upon, the latter being outside the scope of this work. GBM generated the best overall models, followed by Random Forests and the remaining algorithms.

[6] United States Dollar
[7] fred.stlouisfed.org

Algorithm 5 Model tuning algorithm.

Split dataset into training (trs), validation (vls) and testing datasets
for i = 1 to n **do**
    $\text{trs}'_n \leftarrow$ bootstrap resample of trs
    $\text{vls}'_n \leftarrow$ bootstrap resample of vls
    Train a Random Forest model, RF, on trs'_n, with *ntrees* = 500, *max_depth* = 3
    $\text{FI}_n \leftarrow$ RF's feature importance
    for *hc* **in** hyperparameter combinations **do**
        L $\leftarrow$ list of all features
        while L is not empty **do**
            Fit model on trs'_n, using the features in L, the hyperparameters described in *hc*
            Use said model to predict vls'_n
            Remove the least important feature from L, according to FI_n
Average vls' prediction performance for each model, feature subset and *hc*
Train final model on entire trs + vls, with best model features and *hc* of the best model
Use final model to predict the testing set

The features are ranked according to a Random Forest model with small maximum depth and a large number of trees. The maximum depth controls how many splits can occur in a given tree, and therefore how many features are used. It is important to size the maximum depth correctly, since a larger maximum depth would allow a larger percentage of trees to be dominated by strong features, suppressing the importance of other features. However, the depth must not be so small as to inhibit meaningful interaction between features and the expression of their overall importance. The use of a large number of trees follows the law of large numbers, that is, the larger the number of trees employed, the closer the resulting feature importance will be to the expected value. Nevertheless, a sensible number must be chosen since training a larger number of trees requires more computation.

After analyzing the behavior of model α while embedded in the PSPS and its scenario evaluation results, its extreme lack of sensitivity became clear, which proved to be critical for quality of feedback given to the user during the planning and simulation process, as will be shown in section 5.4.

4.5.4 Model β

It is in the context of model α's lack of sensitivity that model β came to be developed, in iterative fashion, with particular focus on the generated scatter plot of the predicted versus the actual values of promotional multiplier for the various data sets. From model α to model β, only the algorithm (GBM) and dataset splits were kept the same, with the automatic feature selection algorithm having been put aside. For both α- and β-type models, the training set spans from the start of 2015 until the end of 2018, the validation set encompasses the whole of 2019, with the testing set spanning 2020. All β-type models are GBMs, with a maximum depth of 5 and a number of trees equal to 50, hyperparameters which came from expert advice.

From the outset, a number of features were considered indispensable, namely, the discount percentage and the month of the promotion at hand, as well as the segment, brand, promotional price and capacity of the EAN in question. Inversely, the average daily sales feature was identified as damaging to the model's sensitivity, as it would overshadow the other features in importance upon its addition. Moreover, the external variables were discarded, since their perceived added value did not compensate for the extra complexity required, given that they would need to be updated on a monthly basis.

When the SR runs the models, the resulting estimated promotional multipliers are capped at a minimum of 0.5 and a maximum of 10, in order to avoid possible outlier effects, as mentioned in section 4.4.

Chapter 5

Results

In this section, the various models' results and metrics are analyzed and compared, from which a number of conclusions are drawn about the categories and retailers. The system deployment is discussed as well as how the data available impacted the work.

5.1 A note on the metrics used

Metrics were generated for both EAN-level and promotion-level predictions, specifically the mean absolute deviation (MAD), the coefficient of determination (R^2), the mean absolute percentage error (MAPE) and bias (BIAS), as well as the weighted version of the latter two, all of which were introduced in section 2.3.5. The weighted metrics use different weights depending on the predictions being evaluated, specifically, for EAN-level metrics the *Average daily sales* of each EAN, $\bar{L}_{e,p}$, is used while for promotion-level metrics the sum total *Average daily sales* of the EANs involved in the promotion, $\bar{L}_p$, is used.

5.2 Model α feature selection

The feature selection results for α-type models are shown in Table 5.1. Here it is clear that the *Average daily sales* feature dominates every model, except for the olive oil model for retailer A where the *Weight of the promotion on the category* was found by the model to be more significant. This is not surprising, as the average daily sales serves as an excellent starting point for estimating how much a given EAN will sell in the next promotion. However, in practice it appears to have a limiting effect on the model's output range, as shown in the scatter plots of Fig. 5.1. Many of the company promotion features are overshadowed by others that encapsulate the promotional context more meaningfully and are more easily comprehended by the model. It is apparent how model α overall gives no weight to key features, namely, the month and the segment. The *Discount*, *Promotional price* and *Promotional price per liter* features are somewhat correlated, such that model α naturally splits the importance among them. The feature importances of α-type models also show us that not a single competitor promotion feature was used in a significant way, possibly

	Olive oil				Vegetable Oil			
	Retailer A		Retailer B		Retailer A		Retailer B	
Feature importance	%	#	%	#	%	#	%	#
Average daily sales	37.28	2	76.93	1	77.95	1	84.34	1
Brand	-	-	-	-	-	-	-	-
Cannibalization	-	-	2.91	5	1.98	6	-	-
Capacity	-	-	0.04	16	-	-	-	-
Commodity monthly percentage price change	-	-	0.41	15	-	-	-	-
Commodity monthly price	-	-	1.73	7	-	-	-	-
Competitor promotional intensity	-	-	-	-	-	-	-	-
Days elapsed since last competitor promotion	-	-	-	-	-	-	-	-
Days elapsed since last promotion	-	-	-	-	-	-	-	-
Discount	-	-	0.76	9	-	-	-	-
Duration	-	-	0.44	13	-	-	-	-
MSRP	-	-	0.74	10	-	-	0.93	7
Month	-	-	0.03	17	-	-	-	-
Number of concurrent promotions	-	-	0.62	11	-	-	-	-
Number of recent competitor promotions	-	-	-	-	-	-	-	-
Number of recent promotions	-	-	-	-	-	-	-	-
Percentage of EANs included in the promotion	5.17	3	0.00	18	-	-	0.36	9
Promotional intensity of an EAN	3.88	5	3.70	3	3.29	4	2.18	5
Promotional intensity of the category	-	-	1.25	8	-	-	3.66	3
Promotional price	-	-	1.89	6	-	-	1.44	6
Promotional price per liter	-	-	4.54	2	4.89	3	0.78	8
Recent competitor promotional intensity	-	-	-	-	-	-	-	-
Segment	-	-	0.43	14	-	-	-	-
Weight of an EAN on the promotion	49.05	1	2.97	4	9.14	2	3.96	2
Weight of the promotion on the category	4.62	4	0.61	12	2.75	5	2.35	4

Table 5.1: Feature importances for the various α-type models.

indicating to a grave shortcoming in the way the effect of the competition was captured. Alongside the feature selection, an exhaustive grid search was performed for each model fitted, as described in Algorithm 5, on the two hyperparameters mentioned in section 4.5.3, namely, for a *max_depth* of 3, 4, 5 and 6, and for *ntrees* of 50, 60, 70, 80, 90 and 100. The hyperparameters selected are shown in Table 5.2, where a consensus of a maximum depth of 3 was reached, with the number of trees varying for the various models.

5.3 Model β feature selection

As mentioned in section 4.5.4, unlike model α, the features were selected iteratively and based of expert knowledge, starting with a couple of nonnegotiable ones, namely, the *Discount*, *Month*, *Segment*, *Promotional price*, *Capacity* and *Brand*, the latter only for the vegetable oil category, since the manufacturer has more than one prominent vegetable oil name brand in the market in question. From this base group of features, the rest were added and reasoned about, comparing their effect on training and validation metrics, and a few were ultimately selected, except for the *Average daily sales* feature which was removed from consideration. The final iteration results are shown in Table 5.4.

	Olive oil		Vegetable oil	
Hyperparameters	Retailer A	Retailer B	Retailer A	Retailer B
max_depth	3	3	3	3
ntrees	70	80	100	70

Table 5.2: Hyperparameters selected for the various α-type models.

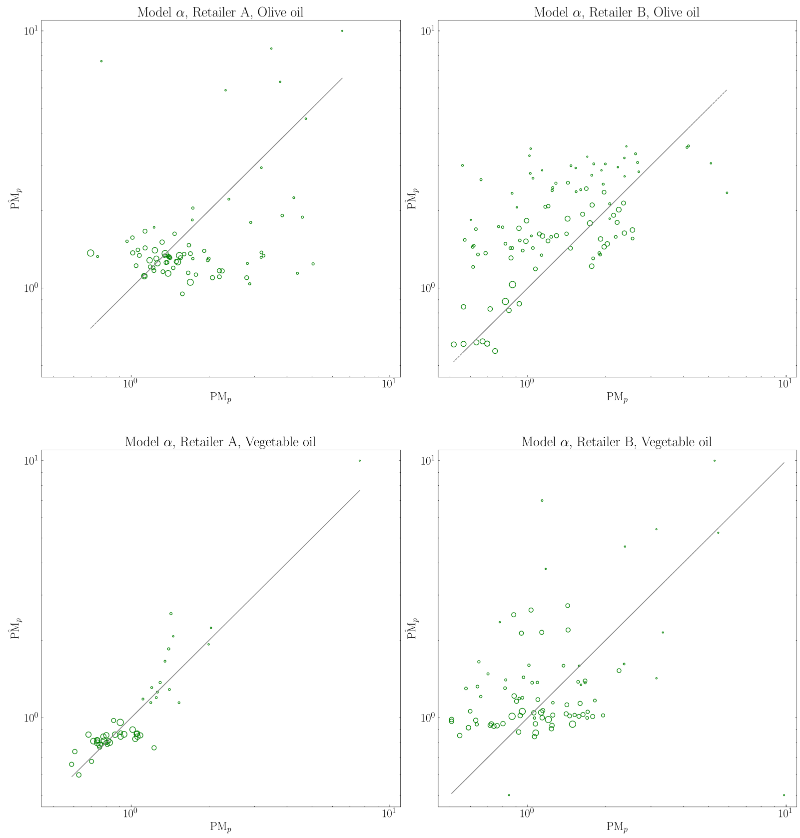

Figure 5.1: Promotion-level α-type model scatter plots for both categories and retailers, showing the relationship between actual versus predicted values of promotional multiplier.

		Olive oil		Vegetable oil	
	Metrics	Retailer A	Retailer B	Retailer A	Retailer B
EAN-level	MAD	0.36	0.39	0.53	1.37
	MAPE	208.38	91.57	301.91	310.59
	BIAS	187.15	70.72	289.06	281.88
	R^2	69.56	68.05	92.07	76.66
	WMAPE	34.69	53.99	15.28	44.72
	WBIAS	0.68	36.52	1.59	22.51
Promotion-level	MAD	0.87	0.73	0.19	0.76
	MAPE	43.80	62.78	12.77	54.84
	BIAS	6.87	51.38	3.30	31.12
	R^2	-77.78	-8.62	82.90	-43.58
	WMAPE	26.34	38.65	10.88	41.34
	WBIAS	-4.48	25.30	0.16	19.99

Table 5.3: Metrics for the various α-type models.

	Olive oil				Vegetable Oil			
	Retailer A		Retailer B		Retailer A		Retailer B	
Feature importance	%	#	%	#	%	#	%	#
Average daily sales	-	-	-	-	-	-	-	-
Brand	-	-	-	-	0.05	9	0.02	9
Cannibalization	-	-	-	-	18.69	2	5.81	4
Capacity	1.76	6	12.47	2	1.88	6	1.91	6
Commodity monthly percentage price change	-	-	-	-	-	-	-	-
Commodity monthly price	-	-	-	-	-	-	-	-
Competitor promotional intensity	-	-	-	-	-	-	-	-
Days elapsed since last competitor promotion	-	-	-	-	-	-	-	-
Days elapsed since last promotion	-	-	-	-	-	-	-	-
Discount	1.19	7	1.57	7	0.65	7	0.49	8
Duration	-	-	-	-	-	-	-	-
MSRP	-	-	-	-	-	-	-	-
Month	1.05	8	0.92	8	0.42	8	1.47	7
Number of concurrent promotions	-	-	-	-	-	-	-	-
Number of recent competitor promotions	-	-	-	-	-	-	-	-
Number of recent promotions	-	-	-	-	-	-	-	-
Percentage of EANs included in the promotion	-	-	-	-	-	-	-	-
Promotional intensity of an EAN	74.42	1	65.57	1	71.38	1	68.84	1
Promotional intensity of the category	2.32	5	3.62	5	-	-	-	-
Promotional price	3.91	4	5.16	4	1.92	5	3.04	5
Promotional price per liter	-	-	-	-	-	-	-	-
Recent competitor promotional intensity	-	-	-	-	-	-	-	-
Segment	5.81	3	7.47	3	2.15	4	8.37	3
Weight of an EAN on the promotion	-	-	-	-	-	-	-	-
Weight of the promotion on the category	9.54	2	3.23	6	2.86	3	10.05	2

Table 5.4: Feature importances for the various β-type models.

5.4 Comparing model results

Comparing the metrics for both types of models, model β comes out as the superior alternative, in all but a few metrics in some specific models, as seen in Table 5.5. Overall, β-type models reveal better performance on the testing dataset, possess stronger R^2 values and, these both being correlated, better scatter plots, when compared to their counterpart, as shown in Fig. 5.2. The α-type models tend to mostly predict values around 1, meaning the model predicts the overwhelming majority of promotions to be close to average, which translates into poor feedback for the planning process. β-type models are in practice more responsive and fit the whole response range better. Although α-type models are generally not far behind their counterparts when it comes to promotion-level metrics, their ineptitude is apparent at the EAN-level, as evidenced by their EAN-level MAPE metrics, which are inflated due to their inadequate modeling of low-selling products, indicated by comparison with the far more redeeming WMAPE metric, which places more focus on high-selling products. It is then immediate that β-type models capture both ends of the spectrum in a superior way, which is also visible in the EAN-level scatters, which can be seen in Appendix A.

5.5 Comparing categories

Of both categories, the vegetable oil category, should, theoretically, be easier to predict, given the weaker competition, as suggested by the commanding market share held by the company, which translates into less relevant competition effects, effects which are hard to capture presently with the data available. This is a possible reason why the *Cannibalization* feature takes on significance in retailer A's vegetable oil β-type model, since in the presence of weak competitor products, the

		Olive oil				Vegetable oil			
		Retailer A		Retailer B		Retailer A		Retailer B	
	Metrics	α	β	α	β	α	β	α	β
EAN-level	MAD	0.36	0.20	0.39	0.28	0.53	0.39	1.37	1.08
	MAPE	208.38	55.68	91.57	48.98	301.91	40.78	310.59	52.50
	BIAS	187.15	24.61	70.72	29.84	289.06	31.20	281.88	33.33
	R^2	69.56	90.00	68.05	78.80	92.07	95.11	76.66	68.18
	WMAPE	34.69	21.45	53.99	37.52	15.28	11.16	44.72	42.68
	WBIAS	0.68	3.01	36.52	22.75	1.59	0.67	22.51	30.40
Promotion-level	MAD	0.87	0.37	0.73	0.37	0.19	0.15	0.76	0.44
	MAPE	43.80	16.92	62.78	29.22	12.77	10.37	54.84	34.49
	BIAS	6.87	-0.70	51.38	16.84	3.30	-0.45	31.12	21.04
	R^2	-77.78	70.15	-8.62	72.40	82.90	85.27	-43.58	22.82
	WMAPE	26.34	16.09	38.65	25.68	10.88	9.04	41.34	35.46
	WBIAS	-4.48	-0.78	25.30	14.95	0.16	-0.03	19.99	26.55

Table 5.5: Metrics for the various α- and β-type models.

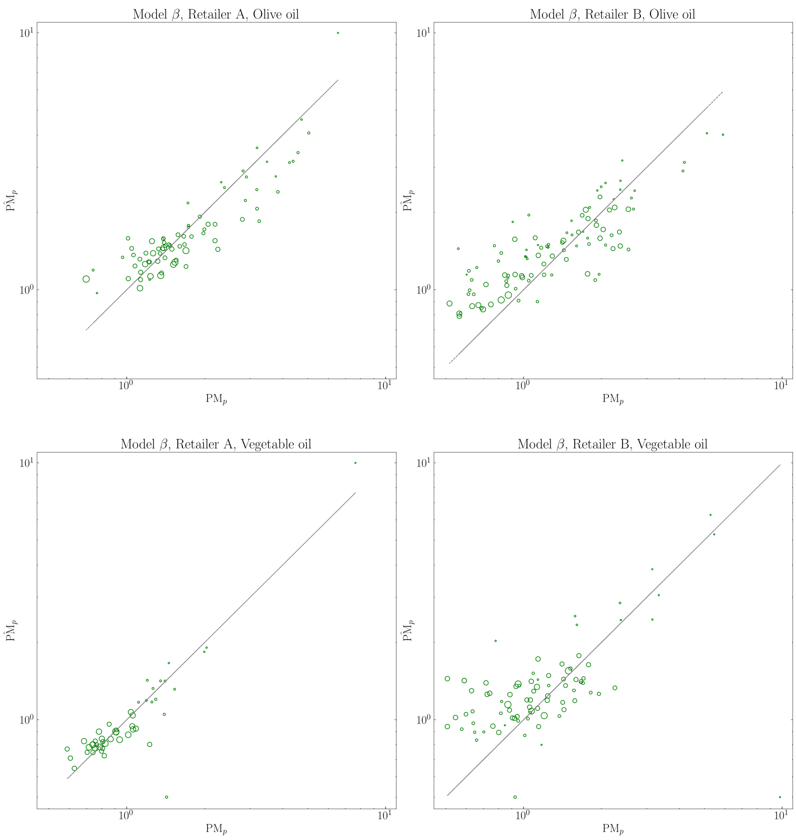

Figure 5.2: Promotion-level β-type model scatter plots for both categories and retailers, showing the relationship between actual versus predicted values of promotional multiplier.

fiercest competition available is found in the company's other products. For retailer B, this is not immediately apparent as both model types struggled to capture *Cannibalization*'s significance, possibly due to the sell-in noise, non-apparent differences between the retailers or a combination of factors.

The olive oil category, being more heavily contested, is naturally dependent on the actions of the competition. However, no feature adequately encapsulated this effect, as mentioned above. This was also noticed experimentally, during the selection of features for β-type models, in which none of the features added value in a compelling way to the model. It is possible that the downside of lacking a feature that adequately models the competitors' effect is softened by the commercial team's expertise in drafting promotional plans. Retailers often relay onto them when the competition is planning to promote their products such that the team is able to plan accordingly.

5.6 Comparing retailers

Retailer B is harder to predict than retailer A, despite the former having a larger dataset. This could be attributed to differences between the retailers, specifically differences in the products offered for each category. Fig. 5.3 shows the Pareto[1] distribution for the company's products in terms of liters sold, demonstrating the wider variety of products offered by retailer B in each category, one of the traits the retailer is known for.

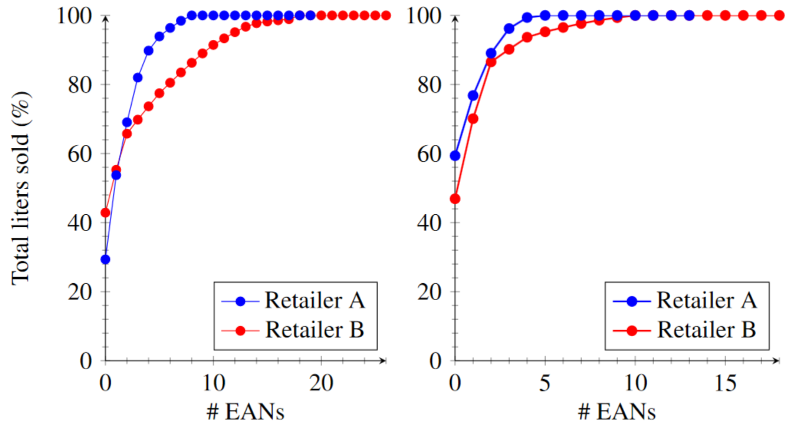

Figure 5.3: Pareto distribution of liters sold by the company, for both retailers and categories (Olive oil on the left, vegetable oil on the right).

However, a stronger justification is the fact that retailer B's models are trained on a sell-in approximation of the sell-out numbers, which introduces noise in the data that naturally degrades the models' performance. To validate such a hypothesis, given that sell-in data is naturally available to all retailers, the sell-in version of retailer A, referred to as retailer A*, was used to train β-type

[1]Vilfredo Pareto, another famous Italian polymath, coined the Pareto principle after observing that 80% of Italian land was owned by 20% of its population. The principle essentially states that a minority of agents bring about a majority of consequences.

models in order to judge the effects of the sell-in approximation. The metrics of the β-type retailer A* models are shown in Table 5.6, supporting the hypothesis that a large portion of the difference between retailer A and retailer B's model performance is explained by the difference between the quality of their sales data.

		Olive oil		Vegetable oil	
	Metrics	Retailer A	Retailer A*	Retailer A	Retailer A*
EAN-level	MAD	0.20	0.45	0.39	2.44
	MAPE	55.68	72.40	40.78	106.50
	BIAS	24.61	4.17	31.20	24.28
	R^2	90.00	46.94	95.11	-10.10
	WMAPE	21.45	51.59	11.16	77.61
	WBIAS	3.01	-23.94	0.67	-62.15
Promotion-level	MAD	0.37	0.66	0.15	0.76
	MAPE	16.92	39.68	10.37	54.63
	BIAS	-0.70	-19.12	-0.45	-17.24
	R^2	70.15	-21.83	85.27	17.39
	WMAPE	16.09	32.20	9.04	44.00
	WBIAS	-0.78	-22.86	-0.03	-39.84

Table 5.6: β-type model metrics for retailer A and retailer A*.

5.7 System deployment results

The full decision support system was deployed successfully and is, at the time of writing, being actively used by the manufacturer's commercial team. The β-type models were delivered alongside the rest of the tools, for both retailer A and B, for each category combination. Additionally, two extra β-type models were trained to enable the simulation of promotional plans drafted for the smaller retailers, one for each category. These models were trained using every small retailers' sell-in data, as well as retailer B's data, to offset for their small record count. The models essentially behave as if the smaller retailers were retailer B, such that not much confidence has been placed in them, thus the commercial team has been cautioned to take their estimates with caution. The commercial team has spoken out in a fairly positive way about the work done, which they state has improved their promotional plan registry process and streamlined their planning process.

5.8 Comments on data-related impacts

The sales data available to the manufacturer pales in comparison to the wealth of information the large retailer possesses, which involve scanner-level data, often going household-level deep, in the case of retailers with extensive and successful loyalty programs. Such data, besides allowing its holder to understand which products were sold when and at what price point, enables tracking consumer purchase patterns. Furthermore, it helps remove some uncertainty over the promotional plan's execution and reach, since sales could be directly linked to a particular promotion, especially for promotions of restricted geographic coverage. It is not hard to envision using such data to predict stockpiling effects as well as timing post-stocking demand. The manufacturer that gains access to this kind of data can potentially have exceptional insight, that can therefore be leveraged in negotiating future trade promotions, if a sophisticated system of data collection and analysis has been implemented for it.

Chapter 6

Conclusions and future work

The work aimed to evaluate trade promotions from the point of view of a consumer packaged goods manufacturer, with limited and unpolished data, in order to enable their commercial team to avoid below-average promotions and empower them to better negotiate trade promotions with the retailers, especially given the current frequency with which promotions are being held. To this effect, a comprehensive decision support system was developed, to allow the team to plan and simulate specific promotional plans, receiving an estimate of the resulting sales, which then can be used as intelligence for negotiation with the retailers. This work stresses the importance of having access to quality data, which most consumer packaged goods companies do not have. By acting as middlemen to the manufacturers, retailers are granted access to a wealth of data pertaining to the customer base of the manufacturer, as well as their purchase patterns, which could be of great use to the manufacturer. Of the retailers that deal with the manufacturer, the one that provided sell-out data fueled the best models obtained, testifying to the importance of such data's availability. Both parties ought to come together and share information with one another, in order to improve their competitiveness and strengthen their partnership. A tighter manufacturer-retailer relationship would mean more efficient management of inventory and stronger margins for both.

The literature covering the forecasting of sales induced by trade promotions from the manufacturer's perspective is rather scarce when compared to the literature regarding the concerns of the retailer. This comes naturally as a consequence of the retailers' enormous push for research and development, retailers which face fierce competition and are in great need of leveraging the enormous amount of data they possess into competitive advantage, funding many endeavors in the field of sales forecasting and inventory management. In that aspect, this work adds to the literature and helps to ease the gap in it.

An interesting avenue not explored by this work lies in more effectively categorizing and distinguishing products to improve EAN-level predictions and to also allow the methodology to be extended, covering manufacturers with wider and more diversified portfolios of products in one category. This work explored predicting sales for a given product-promotion pair, such that exploring different targets, at diverse granularities, and their trade-offs could prove to be an interesting future research avenue.

References

Abolghasemi, M., Beh, E., Tarr, G., and Gerlach, R. (2020). Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion. *Computers & Industrial Engineering*, 142:106380.

Abraham, M. M. and Lodish, L. M. (1987). Promoter: An automated promotion evaluation system. *Marketing Science*, 6(2):101–123.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

Blattberg, R. C., Kim, B., and Ye, J. (1996). Defining baseline sales in a competitive environment. *Seoul Journal of Business*, 2.

Blattberg, R. C. and Levin, A. (1987). Modelling the effectiveness and profitability of trade promotions. *Marketing Science*, 6(2):124–146.

Blattberg, R. C. and Neslin, S. A. (1989). Sales promotion: The long and the short of it. *Marketing letters*, 1(1):81–97.

Bose, I. and Mahapatra, R. K. (2001). Business data mining—a machine learning perspective. *Information & management*, 39(3):211–225.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

Broderick, A. and Pickton, D. (2005). *Integrated marketing communications*. Pearson Education UK.

Cauchy, A. (1847). Méthode générale pour la résolution des systemes d'équations simultanées. *Comp. Rend. Sci. Paris*, 25(1847):536–538.

Chambers, J. C., Mullick, S. K., and Smith, D. D. (1971). How to Choose the Right Forecasting Technique. *Harvard Business Review*. Section: Forecasting.

Cooper, L. G., Baron, P., Levy, W., Swisher, M., and Gogos, P. (1999). PromoCast™: A New Forecasting Method for Promotion Planning. *Marketing Science*, 18(3):301–316.

Cryer, J. D. (1986). *Time series analysis*, volume 286. Springer.

Devore, J. L. (2008). Probability and statistics for engineering and the sciences.

Divakar, S., Ratchford, B. T., and Shankar, V. (2005). Practice Prize Article— *CHAN4CAST* : A Multichannel, Multiregion Sales Forecasting Model and Decision Support System for Consumer Packaged Goods. *Marketing Science*, 24(3):334–350.

Farris, P. W., Bendle, N., Pfeifer, P. E., and Reibstein, D. (2010). *Marketing metrics: The definitive guide to measuring marketing performance*. Pearson Education.

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5):1189–1232. Publisher: Institute of Mathematical Statistics.

Greenlaw, S. A. and Shapiro, D. (2017). Principles of microeconomics 2e.

Hand, D. J. and Adams, N. M. (2014). Data mining. *Wiley StatsRef: Statistics Reference Online*, pages 1–7.

Hill, R. (2016). *Promotions: do you know what you don't know?* Kantar Worldpanel.

Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

Jensen, R. T. and Miller, N. H. (2007). Giffen behavior: Theory and evidence. Working Paper 13243, National Bureau of Economic Research.

Kuhn, M. and Johnson, K. (2019). *Feature engineering and selection: A practical approach for predictive models*. CRC Press.

Nielsen (2020). 3,5 mil milhões de euros na "Selva Promocional" dos bens de grande consumo – Nielsen.

Stobierski, T. (2020). Willingness to Pay: What It Is & How to Calculate. *Harvard Business School*.

Vitasek, V. (2006). Terms and glossary. *Supply Chain Visions*.

Zorbas, C., Gilham, B., Boelsen-Robinson, T., Blake, M. R., Peeters, A., Cameron, A. J., Wu, J. H., and Backholer, K. (2019). The frequency and magnitude of price-promoted beverages available for sale in australian supermarkets. *Australian and New Zealand journal of public health*, 43(4):346–351.

Appendix A

Model scatter plots

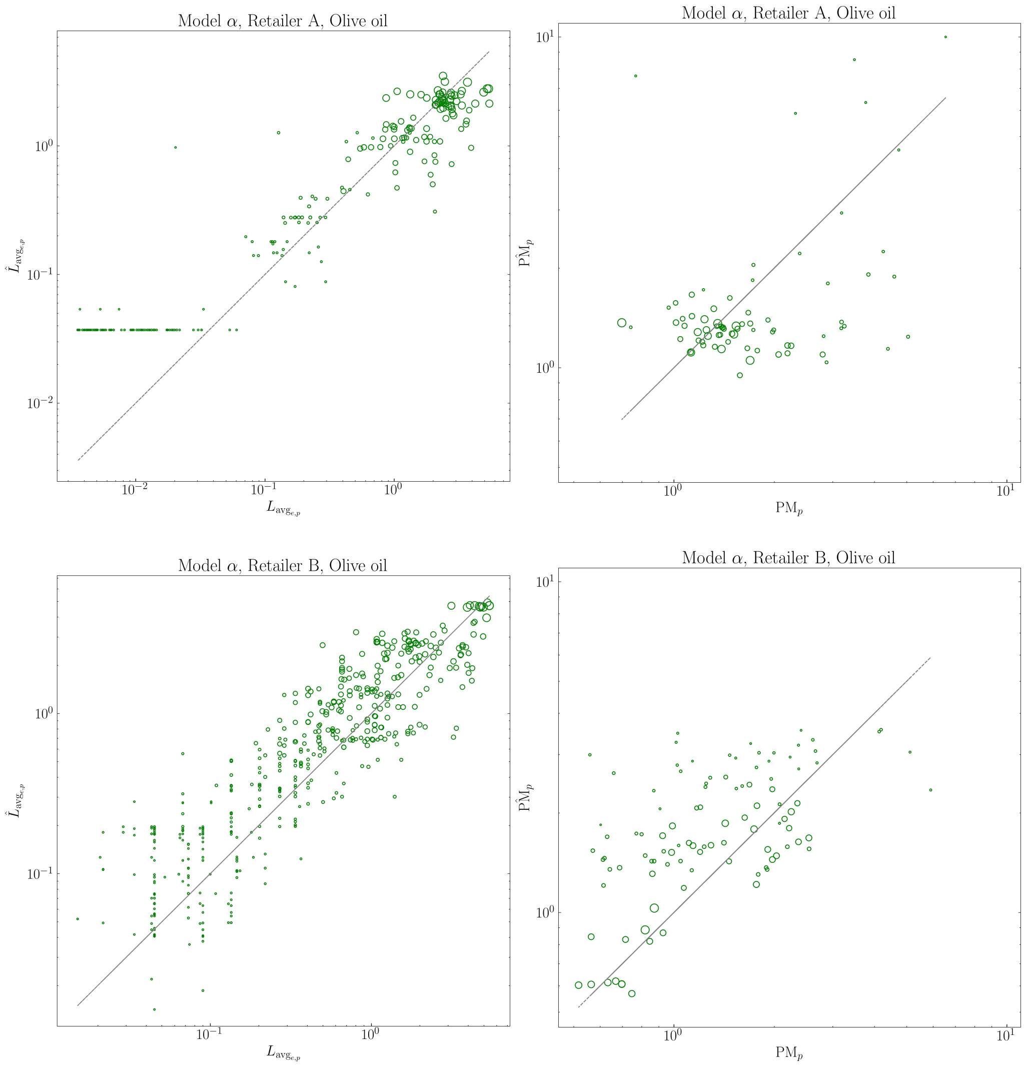

Figure A.1: Olive oil scatter plots. α- and β-type scatter plots on the left and right half, respectively, for retailer A. EAN-level and promotion-level scatter plots are laid side-by-side.

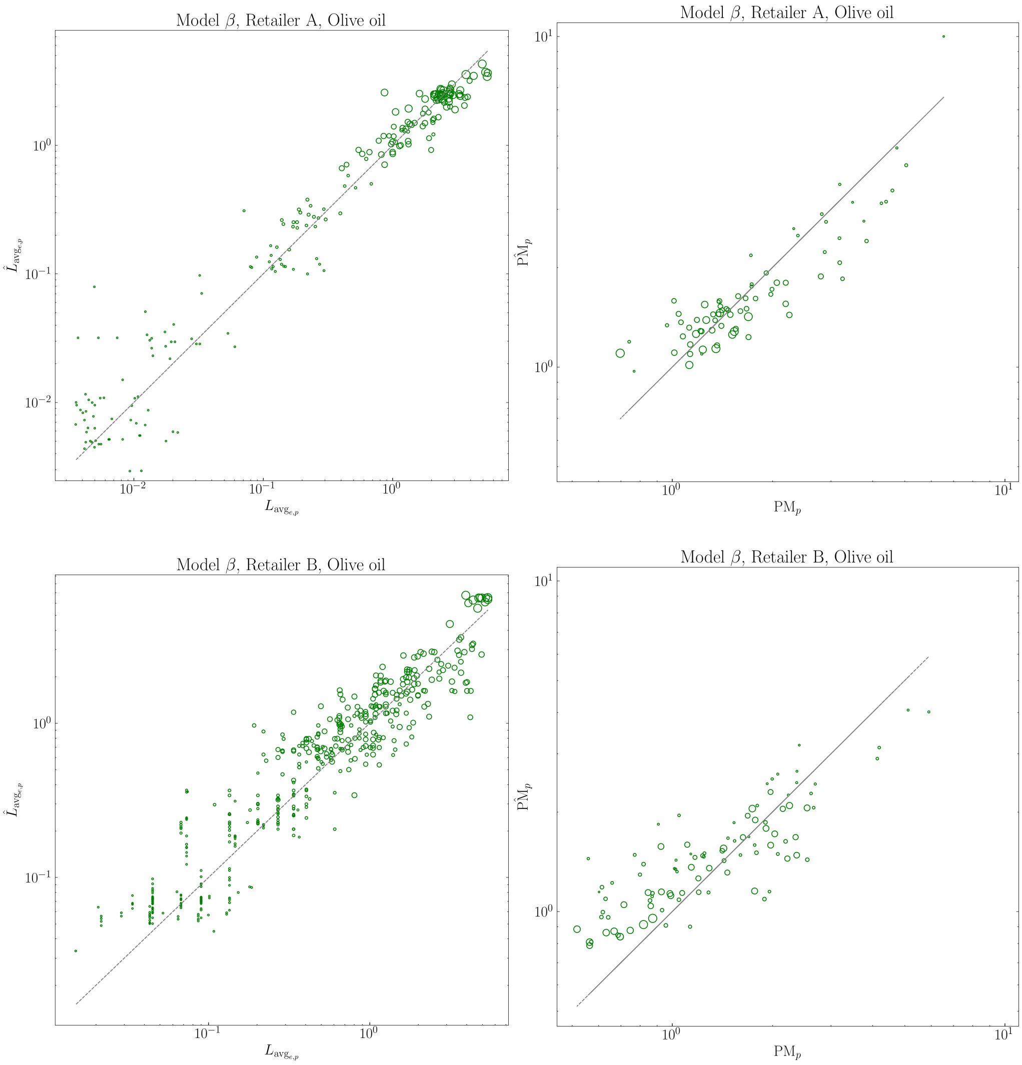

Figure A.2: Olive oil scatter plots. α- and β-type scatter plots on the left and right half, respectively, for retailer B. EAN-level and promotion-level scatter plots are laid side-by-side.

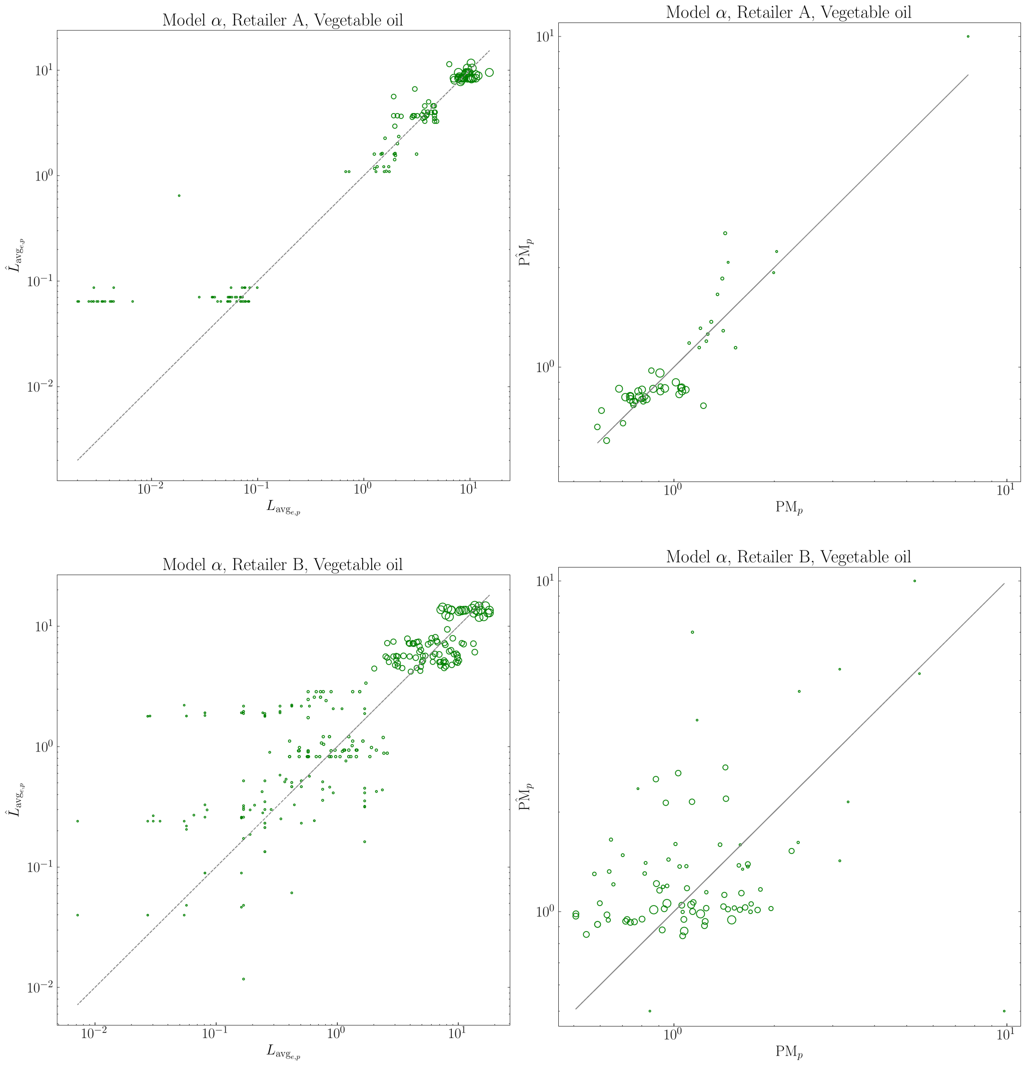

Figure A.3: Vegetable oil scatter plots. α- and β-type scatter plots on the left and right half, respectively, for retailer A. EAN-level and promotion-level scatter plots are laid side-by-side.

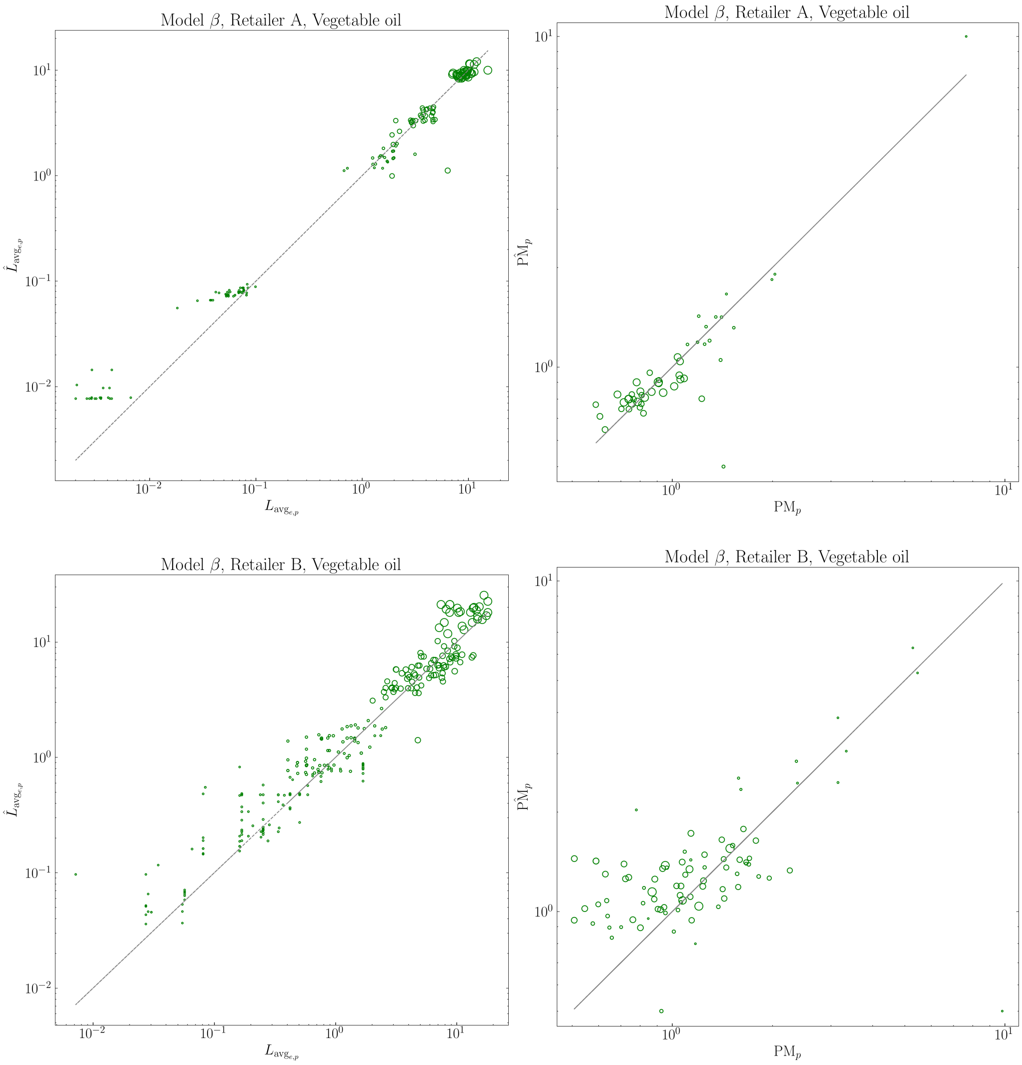

Figure A.4: Vegetable oil scatter plots. α- and β-type scatter plots on the left and right half, respectively, for retailer B. EAN-level and promotion-level scatter plots are laid side-by-side.