Aalto University
School of Science
Degree programme in Engineering Physics and Mathematics

Evaluating cannibalization between items in retail promotions

Bachelor's thesis
October 15, 2018

Olli Herrala

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of bachelor's thesis

Author Olli Herrala	
Title of thesis Evaluating cannibalization between items in retail promotions	
Degree programme Engineering Physics and Mathematics	
Major Systems Analysis	**Code of major** F3010
Supervisor Prof. Fabricio Oliveira	
Thesis advisor(s) D.Sc. Mikko Ervasti	
Date 15.10.2018	**Number of pages** 22 **Language** English

Abstract

In today's competitive retail landscape, promotions are widely used to direct consumer choice and drive traffic and sales. In order to understand the real impact of a promotion, it needs to be decomposed into clear components applicable in decision making. The component we chose for analysis is cannibalization, or how much of the promotion uplift is diverted from the sales of substitute products. This thesis aims to develop a machine learning method for evaluating the extent of cannibalization between individual items from time series sales data.

Cannibalization was determined from sales data as the ratio between the volume drop of the cannibalized product and the volume uplift of the promoted product. Volume was used instead of turnover because of the clearer connection to consumer choice and demand substitution. The method used was an elastic net regularized alternating least squares optimization.

When testing the method on simulated data for three years, we found that the method is stable in that it converges in the same solution independent of the initial guess. The accuracy was found to decrease as the number of products or the noise in the data increased. The method was found to perform better both with an improved baseline model and a longer time window. The running times of the method were reasonably low, and by properly parallelising the calculations, significant further improvements could be easily achieved. The developed method is still rather simple and leaves many open questions for future work. However, even in this form, the method is sufficient for providing estimates that hardly appear in prior literature.

The results of this method are somewhat sensitive to the quality of the data and would likely be more inaccurate with actual sales data from retailers, as the consumer behaviour doesn't follow the assumptions as strictly as in the simulated data set. However, the method does have clear applicability in retail promotion planning, as it nevertheless provides magnitude estimates for individual item pairs, allowing managers to quickly see which products are the biggest cannibals. On the other hand, the method also gives estimates for complementarity, the inverse effect of cannibalization. Overall, promotion planning has large potential in increasing promotion margins and giving companies the competitive edge.

Keywords retail promotions, cannibalization, consumer choice, elastic net regularisation, alternating least squares

Aalto-yliopisto
Perustieteiden korkeakoulu

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi

Tekniikan kandidaatintyön tiivistelmä

Tekijä Olli Herrala

Työn nimi Evaluating cannibalization between items in retail promotions

Koulutusohjelma Teknillinen fysiikka ja matematiikka

Pääaine Matematiikka ja systeemitieteet **Pääaineen koodi** F3010

Vastuuopettaja Prof. Fabricio Oliveira

Työn ohjaaja(t) TkT Mikko Ervasti

Päivämäärä 15.10.2018 **Sivumäärä** 22 **Kieli** Englanti

Tiivistelmä

Nykypäivän kilpailuhenkisessä vähittäiskauppaympäristössä promootioita käytetään laajalti ohjaamaan kuluttajan valintaa ja lisäämään liikennettä ja myyntejä. Jotta promootion todellisen vaikutuksen voisi ymmärtää, se on hajotettava selkeisiin päätöksenteossa hyödynnettäviin komponentteihin. Analysoitavaksi valitsemamme komponentti on kannibalisaatio eli se, miten suuri osa promootion lisämyynnistä johtuu muista korvaavista tuotteista kääntyneistä myynneistä. Tässä työssä pyritään kehittämään koneoppimismenetelmä yksittäisten tuotteiden välisen kannibalisaation suuruuden arvioimiseen aikasarjamuotoisesta myyntidatasta.

Kannibalisaatio määriteltiin myyntidatasta kannibalisoidun tuotteen myyntimäärän laskun ja promootiotuotteen lisämyynnin välisenä suhteena. Myytyä volyymiä käytettiin euromääräisen myynnin sijaan, koska sillä on selkeämpi yhteys kuluttajan valintaan ja kysynnän korvautuvuuteen. Käytetty menetelmä oli elastinen verkko -regularisaatiota käyttävä vuorottelevan pienimmän neliösumman menetelmä.

Testatessa menetelmää kolmen vuoden simuloidulla datalla havaittiin menetelmän olevan stabiili, sillä se suppenee aina samaan tulokseen alkuarvauksesta riippumatta. Tarkkuuden havaittiin huononevan tuotteiden määrän ja datan kohinaisuuden kasvaessa. Menetelmän havaittiin toimivan paremmin, kun myynnin perustaso määritellään tarkemmin tai kun analysoitavaa aikaikkunaa pidennetään. Ajoajat pysyivät kohtuullisen pieninä, ja kunnollisella rinnakkaistamisella voitaisiin helposti saavuttaa merkittäviä edistysaskeleita. Kehitetty menetelmä on edelleen melko yksinkertainen ja jättää paljon tilaa jatkokehitykselle. Kuitenkin jo nykymuodossaan menetelmä on riittävä kirjallisuudessa hyvin vähän esiintyvien arvioiden tuottamiseen.

Menetelmän tulokset ovat jonkin verran herkkiä datan laadun suhteen ja olisivat luultavasti aidolla myyntidatalla vähemmän tarkkoja, sillä kuluttajien käyttäytyminen ei täysin noudata mallin mukaisia oletuksia. Tästä huolimatta menetelmästä on selkeää hyötyä vähittäiskaupan promootiosuunnittelussa, sillä se kaikesta huolimatta kykenee tuottamaan suuntaa-antavia arvioita siitä, mitkä tuotteet kannibalisoivat eniten muita. Toisaalta menetelmä arvioi samalla myös komplementaarisuutta, kannibalisaation käänteisilmiötä. Kaiken kaikkiaan promootiosuunnittelulla on suurta potentiaalia promootiokatteiden nostamisessa ja yrityksille kilpailuedun saavuttamisessa.

Avainsanat vähittäiskaupan promootiot, kannibalisaatio, kuluttajanvalinta, elastinen verkko -regularisaatio, vuorottainen pienin neliösumma

Contents

1 Introduction

In a price-driven retail landscape, different promotions and discounts play a significant role in directing customer choice. The offer of discounts makes customers more likely to buy the product, and often leads to increased total sales (Srinivasan et al. [2004]). This is what retail companies have usually seen in their market analyses, but the concept of customer choice suggests other factors should also be taken into account when analyzing promotion effectiveness.

Promotions are a key element in retail, accounting for 10 to 45 percent of retailers' total revenues (Goad et al. [2015]). However, only 20 to 60 percent of the promotions succeed in increasing the margins, while others simply don't generate enough sales to be beneficial (Goad et al. [2015]). This is largely due to the retailers not being able to understand or analyze all the components of promotion margin generation (Walters [1991]).

When calculating the incremental margin resulting from a promotion, many different components need to be taken into account. These include stock-up and cannibalization, along with multiple other phenomena. Stock-up refers to consumers buying great amounts of a discounted product so they don't need to buy it later on full price, and cannibalization to when a consumer chooses a promoted product over a similar product they would have otherwise bought. In order to run effective promotions, as many of these components need to be understood as clearly as possible. However, in the context of this thesis, the objectives were narrowed down to one single component. Cannibalization was chosen because of the importance of the effect, and the complexity of the problem.

This thesis has been written as part of an internship at Sellforte Solutions Ltd. where the goal was to develop a method for determining cannibalization between individual item pairs during promotions. The first step is creating simulated data for developing and testing of the methods, and after that, we propose a method for extracting cannibalization information from the data. This method should help retailers make informed decisions about their promotions by providing insight into which products strongly cannibalize each other.

The main requirements that arise from the set goals are reliability and speed. For the method to have any practical applicability, it needs to be accurate enough to provide estimates with some predictive power. It is also preferable that the method can calculate the estimates in a reasonable time on a standard computer. The main application will be predicting promotion performance by learning which promotions take sales away from other products instead of bringing actual new sales.

Section 2 presents background information on cannibalization in retail as a phenomenon, along with an overview of previous research on the topic. Section 3 presents the suggested method and describes the simulated data and the underlying assumptions about consumer choice. Sections 4 and 5 present the results about the goodness of the method, and conclusions about the practical usability.

2 Background

2.1 Cannibalization

In 1976, James Heskett defined cannibalization as "the process by which a new product gains a portion of its sales by diverting them from an existing product" (Heskett [1976]). The definition has since extended to cover other cases of a product diverting sales from another product, and is commonly used today in the context of promotion efficiency analysis.

The theory behind cannibalization lies in consumer theory and substitute goods. In consumer theory, two products are substitutes if a rise in the price of product A causes the demand for product B to rise. Examples of substitutes include butter and margarine (slightly different products that are used for the same purpose), Coca-Cola and Pepsi (competing brands), and ice cream cones and sticks (different form or pack size). This substitutability results in promotion cannibalization. The phenomenon is visualized in Fig. 1, where products A and C are promoted (e.g. discounted) for weeks 6 and 7. This results in a significant volume uplift for the promoted products, but also a simultaneous drop in the sales of the non-promoted product B.

Cannibalization results from customers temporarily switching from other products to the promoted product due to lower price or greater visibility. What this means for the retailer is that part of the uplift in a promoted product comes from the sales of other products, thus causing a decrease in total sales not observable from analyzing the promoted product alone. Cannibalization can also be positive, for instance promoting sausages should increase mustard sales by intuition. This phenomenon is called complementarity, or sometimes halo.

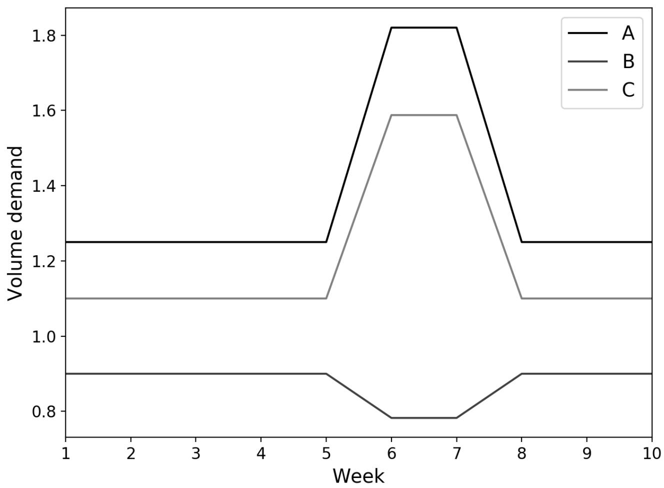

Figure 1: Sales time series for three products A, B and C in weekly resolution.

The total magnitude of the drop in the sales of other products has been approximated to be on average around 30% of the promotion uplift in a set of normal grocery items: canned tuna, tissue, shampoo and peanut butter (Heerde et al. [2002]). This means that 30% of the volume uplift in a promoted product comes from the sales of other products. This is particularly detrimental to the total incremental margin, as the sales shift from normal-priced products to a product that might be heavily discounted.

Multiple publications suggest that the cross elasticity of demand, or the percentage change in the demand of a product resulting from a 1% change in the price of another product, can be assumed roughly constant (Frank [2008]). For substitute products, a decrease in the price of product A results in a decrease in the demand of product B. This can be intuitively explained by the consumers choosing the now cheaper product A over product B. For complementary products, the effect is opposite: the changes in demands of the two products have the same sign. This motivates that cannibalization between two products could be defined from the changes in demand for those products during promotions.

Cannibalization and complementarity can also be estimated by basket analysis. In basket analysis, the goal is to learn association rules A $\rightarrow$ C, where A is an itemset called antecedant and C an itemset called consequent (Agrawal et al. [1993]). For these rules, various metrics can then be derived for describing the goodness of the rule. A common example of an association rule is $\{diapers\} \rightarrow \{beer\}$, meaning that if a person buys diapers, it is likely that he/she will also buy beer. This would be an easy way to get some insight on which products appear or do not appear together, as there are many implementations available. However, these methods are often too computationally demanding, and thus we attempt to implement a method that gives better results faster. Furthermore, basket analysis would be more useful for estimating complementarity than cannibalization, as complementarity can be estimated more directly from receipt data, while cannibalization would require information on which products are not on a given receipt.

2.2 Previous research

The term cannibalization in the context of retail promotions dates back to at least 1972, when a constant fraction was used in calculating the net incremental share effect due to a promotion (Little [1972]). As both data and computational power became more available, approaches using large datasets became possible. Blattberg and Wisniewski found that price competition happens inside so-called price-quality tiers (Blattberg and Wisniewski [1989]), and a product cannibalizing a higher tier is not common. This suggests that customers who buy premium products do not switch to a lower price-quality brand unless there is a significant enough price cut to justify the low quality, while the customers that usually buy the cheap brand are willing to try the premium brand when they can afford it.

Mason and Milne identified pairwise cannibalization for cigarettes using overlapping customer niches calculated from market research data of 9659 observations (Mason and Milne [1994]). Lomax used the deviation from expected sales in measuring cannibalization (Lomax [1996]). This laid foundation for algorithms built on baseline estimates. Srinivasan et al. improved the approach from Lomax by expanding the possibility of cannibalization across different product families (Raghavan Srinivasan et al. [2005]). Cooper showed the asymmetry of cross-brand elasticities, implying that cannibalization inside a category is not constant, but rather some brands or even products are more vulnerable to cannibalization (Cooper [1988]). Finally, in 2009, Yuan et al. calculated pairwise cannibalization, or "diversion ratios", for orange juice category in new product introduction (Yuan et al. [2009]). Their approach was based on first calculating cross-price elasticities, which are then converted to diversion ratios. In 2002, Abere et al. converted volume cannibalization to sales cannibalization simply by multiplying with the unit price ratio between the cannibalized product and cannibalizing product (Abere et al. [2002]).

In summary, the previous research on cannibalization is mostly conceptual, but suggests that sales data could be used to determine estimates for pairwise cannibalization. Additionally, most of the publications on cannibalization strongly emphasize the managerial significance of understanding the phenomenon, underlining the importance of research on the topic. As further reasoning, promotion decomposition and cannibalization have been researched in multiple companies, but their research is confidential. However, they all promise results that can provide great business understanding and increased margins (RELEX [2018], Revionics®, dunnhumby [2015], Goad et al. [2015]).

3 Methods

3.1 Baseline and uplift

Based on the literature on consumer choice theory, as well as intuition, cannibalization was chosen to be determined from volume changes rather than sales (turnover). The main reason for this was that volume uplift behaves in a simpler way, as visualized in a simplified example in Fig. 2. In the example, volume uplift is assumed linear w.r.t. the price index (price of the product on discount scaled to normal price being 1), volume cannibalization is a constant 10%, and the reference level for both volume and turnover is 1. The solid lines represent the sales and volume of a promoted product, and the dots represent cannibalization. What happens with discounts greater than 30% is that the volume uplift is not enough to compensate for the discount, and the turnover starts decreasing with increasing discounts. However, more volume gets intuitively cannibalized as consumers prefer the discounted product even more with a high discount. This leads to the cannibalization estimate exploding as the turnover uplift approaches zero, and eventually turn into a high complementarity as discounts greater than 60% decrease turnover for the promoted product.

In order to extract from the data the demand changes caused by promotions, a baseline for the sales volume, or the number of product units sold, is needed. We define baseline as what the weekly sales volume for the product would have been without any promotion. If we are able to calculate a reliable estimate of the baseline, calculating an estimate for cannibalization becomes possible. Unfortunately, analyzing and decomposing product sales time series is a complicated task where multiple factors need to be taken into account.

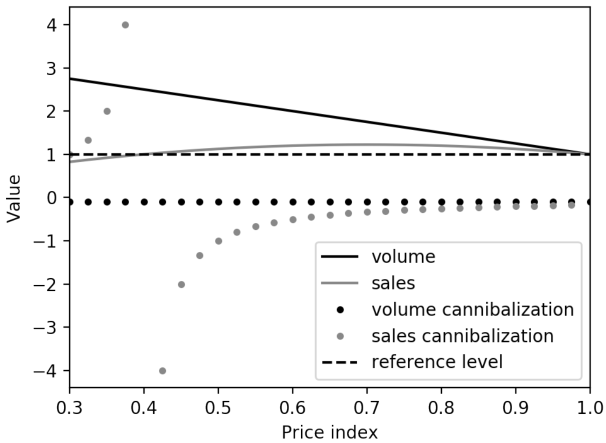

Figure 2: Volume, sales and respective cannibalizations as functions of price index.

One approach to the baselines would be to use a forward naive estimate for the promotion periods. A naive forecast means estimating the value predicted to be the same as the previous reliable observation, $\hat{y}_{t+h|t} = y_t$. This approach is often used in forecasting economic and financial time series, but does not work well on data with a trend or seasonality. A slightly more advanced alternative would be a linear interpolation, where a straight line is fitted between the start and end points of the prediction period. This is clearly better when there is a clear trend in the data, but still fails with high-frequency seasonality. A potential option would be to use a time series decomposition method like Prophet by Facebook (Taylor and Letham [2017]), which is based on an additive model separating trend, seasonality and holiday effects. It is also robust to missing data and outliers, and would clearly be a major improvement to linear interpolation. Nevertheless, such a model was disregarded in this thesis, as the focus is learning cannibalization from measured uplifts.

To avoid dealing with trend and seasonality, we only use the first point of each promotion period and compare it to the last point before the promotion. This reduces the number of data points available for the analysis, but makes it possible to proceed without a more sophisticated baseline model. This is illustrated in Fig. 3, where the solid black line represents actual sales observable from the data, and the dash-dotted line is the base sales that the uplift sales were calculated from. The dashed line shows the baseline estimate. First, there is a linear rising trend for six weeks, and after that the volume stays at around 2.5. We see that the model fails to predict the trend, as was expected for a naive model, but the performance is clearly better in the uplift where there is no trend. A linear interpolation would likely perform better, but as said before, we focus on developing the method for determining cannibalization from a proper data set.

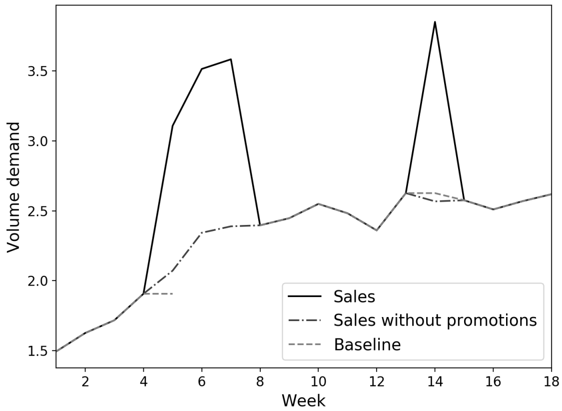

Figure 3: Sales and baseline for a single product in weekly resolution.

We start by calculating the changes in volume demand for each product for the first week of each promotion, creating a $T \times N$ matrix $\boldsymbol{U}$, where T is the number of weeks and N is the number of products in the data. $\boldsymbol{U}_{ti}$ is simply the measured change in demand compared to the baseline for item i on week t. We can then split $\boldsymbol{U}$ into two matrices of the same shape as $\boldsymbol{U}$, one containing uplifts for promoted products and the other containing volume downlifts for cannibalized products. We define cannibalization as the best solution $\boldsymbol{C}$ for the equation

$$\boldsymbol{U'C} = \boldsymbol{D}, \tag{1}$$

where $\boldsymbol{U'}$ is the matrix containing only the promotion uplifts (defined as the volume change for a promoted product) and $\boldsymbol{D}$ is the matrix containing the volume drops for cannibalized non-promoted products. $\boldsymbol{C}_{ij}$ describes how much product i cannibalizes product j, namely the ratio between the volume drop in j caused by i and the volume uplift in i. We define $diag(\boldsymbol{C}) = 0$ because a product does not cannibalize itself. However, $\boldsymbol{C}$ is not required to be symmetric. In Fig. 1, the drop for product B is about 0.12 units, and the uplift for A roughly 0.55 units. If we assume all of B's volume downlift to be caused by A, we get $\boldsymbol{C}_{A,B} = \frac{0.12}{0.55} \approx 0.22$. This definition contains the major assumption that cannibalization is approximately linear with only one parameter for each item pair. In reality, it is possible that different campaigns cause different levels of cannibalization. For example, a big TV ad is likely to bring customers in just to buy the promoted product without thinking about alternatives (very low cannibalization), while an in-store ad causes heavier cannibalization as customers temporarily switch to the promoted product. This could be taken into account by adding another dimension to $\boldsymbol{C}$ for the promotion type, but for now, the method should be used for promotions of a single promotion type.

The least squares solution for this linear regression problem is $\hat{\boldsymbol{C}} = \text{argmin}_{\boldsymbol{C}} \|\boldsymbol{D} - \boldsymbol{U'C}\|^2$. Using a binary promotion matrix $\boldsymbol{P}$, where $\boldsymbol{P}_{ti} = 1$ if product i is promoted on week t and 0 otherwise, $\boldsymbol{U'}$ becomes $\boldsymbol{U} \circ \boldsymbol{P}$, where $\circ$ denotes a Hadamard or elementwise matrix product. On the other hand, $\boldsymbol{D}$ can be expressed as $\boldsymbol{U} - \boldsymbol{U'}$, and thus Eq. 1 becomes

$$(\boldsymbol{U} \circ \boldsymbol{P})\boldsymbol{C} = \boldsymbol{U} - (\boldsymbol{U} \circ \boldsymbol{P}) \ \ (= \boldsymbol{U} \circ \neg \boldsymbol{P}). \tag{2}$$

In order to avoid overfitting the model to the data, regularization is necessary. Furthermore, when there is too little data considering the number of variables fitted, $\boldsymbol{C}$ will have free variables that need to be regularized. We use the elastic net regularization from scikit-learn (Pedregosa et al. [2011]), which combines ridge and lasso regularizations. The elastic net regularized estimate is

$$\begin{aligned} &\hat{\boldsymbol{C}} = \underset{\boldsymbol{C}}{\text{argmin}} \, \|\boldsymbol{D} - \boldsymbol{U'C}\|^2 + \lambda_2 \|\boldsymbol{C}\|^2 + \lambda_1 \|\boldsymbol{C}\|_1 \\ &\text{subject to } diag(\boldsymbol{C}) = 0, \end{aligned} \tag{3}$$

where $\|\boldsymbol{C}\|$ is the Euclidean norm, also known as the 2-norm and $\|\boldsymbol{C}\|_1$ is the 1-norm. λ_2 and λ_1 are used to set the ratio between the two regularizations. A p-norm is defined as $(\sum_{i=1}^{n} (|x_i|^p)^{1/p}$, and is commonly used for determining the length of a vector. The combination of the two norms is useful, as it combines two very desirable properties. The 1-norm from lasso regularization results in a sparse result matrix, while ridge regularization alone tends to estimate a non-zero value for each parameter. The main benefit of ridge regression is that the quadratic penalty makes the loss function strictly convex.

3.2 Self-consistency

Eq. 2 still fails to take simultaneous promotions into account. To illustrate the problem, we consider the dataset of three products A, B and C in Fig. 1. If A and C are put on promotion for a week, the equation for that week t becomes $([U_{t,A}, U_{t,B}, U_{t,C}] \circ [1, 0, 1])\boldsymbol{C} = [U_{t,A}, U_{t,B}, U_{t,C}] \circ [0, 1, 0])$ and

$[U_{t,A}, 0, U_{t,C}]\boldsymbol{C} = [0, U_{t,B}, 0]$. This states that the promotions on A and C can cannibalize the sales of B, which is true. However, for this method to work, we need to take into account that, assuming A and C are substitutes or complements on some level, there can be some cannibalization effect between them. Whether this cannibalization is the same as when they are promoted separately is debatable, but in this work, we assume that the effect is at least similar enough to not cause significant errors in the results. In Fig. 4, the effect of this cross-cannibalization is visualized by the dashed lines.

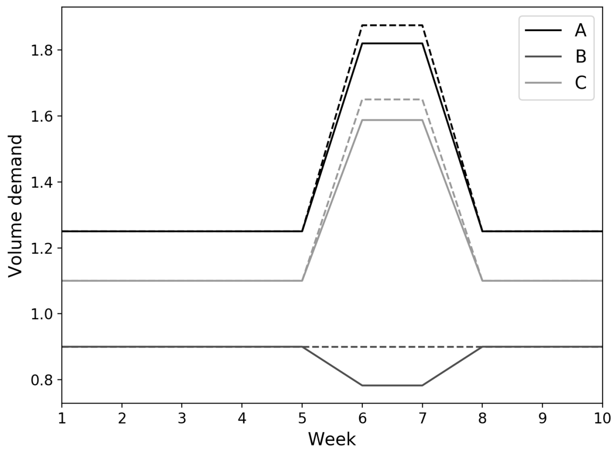

Figure 4: Sales time series for three products A, B and C in weekly resolution. Dashed line represents the demand change without cannibalization, solid line shows the observed demand.

Examining how the cannibalization between promotions behaves is a complicated topic, and thus out of scope. However, without this assumption, we would not be able to use weeks with multiple promotions, which would render this method useless for actual retailer sales data. With this assumption, it is possible to define an equation for the "true" uplifts $\boldsymbol{U'}$ with the effect of cross-cannibalization removed. The true uplift of product A in the example case would be $U'_{t,A} = U_{t,A} - U'_{t,C}C_{C,A}$, and thus

$$\boldsymbol{U'} = \boldsymbol{U} - (\boldsymbol{U'} \circ \boldsymbol{P})\boldsymbol{C}. \tag{4}$$

In an ideal situation with no noise and correct $\boldsymbol{C}$, the promotion matrix would be unnecessary, as $\boldsymbol{U'}$ would get a value of 0 for non-promoted products, such as B in Fig. 4. However, as a result of noise and suboptimal values for $\boldsymbol{C}$ during the calculation, it is necessary to mask the uplifts to be zero where there is no uplift in order to avoid allocating the cannibalization effect to products that are not even promoted.

Because it is now possible for a product to simultaneously have an uplift from a promotion on that product, and a volume drop from other promoted products, it is necessary to constrain the diagonal of $\boldsymbol{C}$ to zero in order to prevent accidentally learning that a product cannibalizes itself. This is done by utilizing the independence of the columns $\boldsymbol{C}_j$, as was done in SLIM (Ning and Karypis [2011]), a similar method in the field of recommender systems. Each column $\boldsymbol{D}_j$ can be calculated separately from $\boldsymbol{U'C}_j = \boldsymbol{D}_j$. By setting $\boldsymbol{U}'_j$ to zero, $\boldsymbol{C}_{jj}$ also becomes zero. What this means in practice is that we simply don't use the uplift of a product to determine the volume drop for that product. In addition to this, the column independence conveniently makes it possible to parallelize the calculations, which greatly increases the applicability to real commercial customer analyses where the number of products is large.

The structure of the problem is as follows: $\boldsymbol{C}$ can be calculated if we know $\boldsymbol{U'}$. However a correct solution for $\boldsymbol{C}$ is needed for determining $\boldsymbol{U'}$. To solve a problem like this, the alternating least squares (ALS) was chosen. In ALS, an optimization problem of two sets of unknowns is solved by alternately fixing one of the (sets of) variables, reducing the problem to a linear regression that can be solved with ordinary linear regression (OLS). In OLS, the result is guaranteed to be optimal (minimal MSE), and thus the accuracy of the solution improves on each iteration until convergence. This is shown in the inner loop of Algorithm 1.

Algorithm 1.

A pseudocode example of the implemented method

```
Split data into training and validation sets
While validation set R^2 larger than on previous iteration:
    While no convergence:
        Update uplifts based on latest cannibalization estimate
        Update cannibalization estimate based on new uplifts
    Multiply λ1 and λ2 by 0.95 to reduce regularization
    Calculate new R^2 for validation set downlifts
```

While the described loop results in a local optimum, and an estimate for $\boldsymbol{C}$, we want the method to have predictive capabilities and avoid overfitting. In overfitting, the results explain the training data well, but fail to predict the values for a validation data. This results from fitting a parameter based on a single outlier, for example. The approach taken for avoiding overfitting in this method was to start with high regularization parameters to keep the cannibalization coefficients constrained. After the iterative algorithm converges at a solution, the coefficient of determination R^2 is calculated for a validation set initially separated from the data and the regularization hyperparameters λ_1 and λ_2 in Eq. 3 are multiplied by 0.95, resulting in more freedom for the cannibalization coefficients. This is repeated until the R^2-value is smaller than on the previous iteration, at which point we conclude that the method is starting to overfit and choose the previous solution as the final result. This is also shown in the outer loop of Algorithm 1. R^2 was chosen as the metric here because of its better comparability between datasets compared to absolute training and validation errors.

Furthermore, regularizing the cross-cannibalization effect even lightly could be a valid addition to Eq. 3, as the effect should not be very large. This would be done by adding a penalty for large promotion cannibalizations, for example $\alpha\|\boldsymbol{P} \circ (\boldsymbol{U} - \boldsymbol{U'})\|^2$. However, this was not added, as it would have required modifying the elastic net implementation.

3.3 Data

There are two main reasons for creating a simulated dataset. First, the applications of these methods are mainly commercial and the data used is therefore sales data from retailers, which is covered by non-disclosure agreements. Thus, in order to validate the method proposed in this thesis, it is necessary to create a realistic nonconfidential dataset. Second, the simulated dataset allows estimating the goodness of the results, as cannibalization values are defined in the simulation, and therefore the real answers are exactly known. This makes it possible to compare the results to known correct values and calculate the errors, in addition to allowing tests on specific features of the model.

The main features required for the dataset are the ability to add any number of products with different baseline sales, adding noise, and adding promotions with cannibalization effects. This allows us to examine the sensitivity of the models with respect to the signal-to-noise ratio and the number of products.

The main assumption in the data is that cannibalization can be presented as a $N \times N$ matrix of constant scalars, where N is the number of products examined. This implies that cannibalization between two products is not dependent on the discount percentage or promotion type. This is supported by the fact that the cross elasticity of demand is often given as a single value for an item pair.

The format that the method was finally tested with had no seasonality, as that should be taken care of by a separate baseline method. The sales volume for each product was set to be one unit, and the volume uplift of a promotion to be 50%. The base sales volume is first distorted by adding Gaussian noise with a zero mean and a given standard deviation or noise. The promotions are set in a repeating pattern of two weeks of promotion and two weeks of regular sales, with a 20% chance of each product being promoted each promotion week.

The cannibalization is applied to the volume data by first creating a matrix $\boldsymbol{C}$ with a mean value of -0.1 and a standard deviation of 0.075, rounded to a precision of 0.05. This is then applied to the volume data according to the assumptions presented with the method.

4 Results

The developed method was applied to a simulated data set with seven products with a noise level of 0.5%, and the resulting matrix $\boldsymbol{C}$ is shown in Table 1. Even though all the values are not rounded to a precision of 0.05 as they should be, the average of the non-diagonal coefficients is -0.098, which is very close to the expected value of -0.1. The mean absolute error is 0.008, which is also relatively small. It can also be seen that most of the values are close to multiples of 0.05. This means that while the method is unable to give perfectly accurate results for individual coefficients, as was expected for noisy data, the results looks promising.

The method implemented in Python was also relatively fast, since calculating the cannibalizations between 40 products for three years of actual customer data from one store could be be done in approximately two hours on a modern laptop (i5-7360U with 8GB of RAM). We found that for the data used, 40 products was just enough to get the R^2 value sufficiently high without using more products than necessary.

4.1 Sensitivity

A test script was created for testing the sensitivity of the method with respect to the number of products (denoted by N) and the level of noise in the data. Cannibalization factors are calculated from simulated data of three years, for 10 randomly generated initial guesses for cannibalization. For all tested

Table 1: Coefficients $\boldsymbol{C}_{i,j}$ from the simulated dataset.

	A	B	C	D	E	F	G
A	0	-0.10	-0.05	-0.08	-0.03	-0.12	-0.16
B	-0.09	0	-0.10	-0.11	0.00	-0.25	-0.11
C	-0.04	-0.14	0	-0.12	0.04	-0.06	0.02
D	-0.20	-0.10	-0.19	0	-0.08	-0.18	0.11
E	-0.16	-0.26	-0.19	0.06	0	-0.19	-0.20
F	0.00	-0.09	-0.05	-0.20	-0.09	0	-0.11
G	0.01	0.00	-0.15	-0.14	-0.10	-0.11	0

combinations of noise and N, the method converged to the same solution in all of the 10 scenarios independent of the initial guess.

An example of the convergence of the coefficients on consequent iterations is seen in Fig. 5, where we have 8 coefficients from a set of six products. The initial guess is seen to be bad, which was expected, as it is generated completely randomly. It can also be seen that in the beginning of the convergence, there is slight oscillation in some of the coefficients. This behavior is in the nature of gradient methods: when the step towards the minimum is too long, it creates oscillation around the optimum. This is why it was necessary to limit the step size by always taking a weighted average of the last two values, weighting the previous result heavily. Without this smoothing, the method could possibly converge faster, but the oscillation would also be greater and it would take longer for it to even out, diminishing the benefit.

After numerically confirming that the method converges in the same optimal solution independent of the initial values, the test is modified so that a new dataset is simulated for each iteration and each combination of noise and N is tested on 100 simulated datasets with the same cannibalization matrix. This allows us to calculate the root mean squared error (RMSE) for each item-item cannibalization factor to get a single metric for measuring the goodness of the method. RMSE was chosen as the performance metric because of relatively good interpretability. The unit of RMSE is the same as for the values it is calculated from. What this metric measures in practice is the difference of the results and known correct values in a very similar way as standard deviation. The smaller the RMSE, the closer the results are to the correct value.

Mean squared error or MSE is defined as the average of the squared difference between estimates $\hat{\boldsymbol{C}}$ and the correct values $\boldsymbol{C}$. This can be formulated as $MSE(\hat{\boldsymbol{C}}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(\hat{\boldsymbol{C}}_{i,j}-\boldsymbol{C}_{i,j})^2}{n^2-n}$, where $i \neq j$. RMSE is then simply derived as $RMSE(\hat{\boldsymbol{C}}) = \sqrt{MSE(\hat{\boldsymbol{C}})}$.

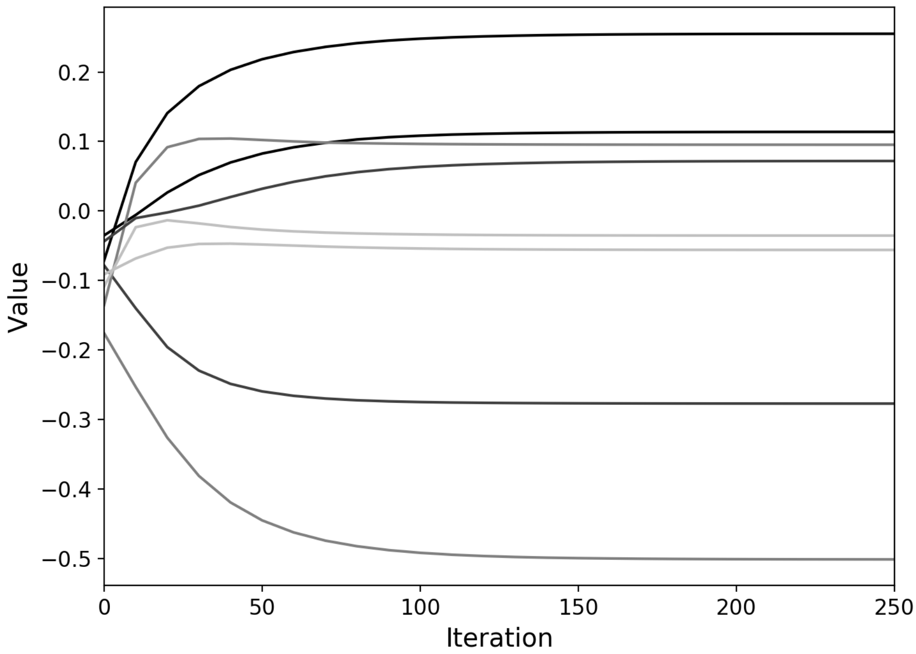

Figure 5: Values of eight $\boldsymbol{C}_{i,j}$ elements at each step of the algorithm iteration.

The test script was run again with 11 different values of N and 9 different noise levels, resulting in 99 test cases in total, each calculated with 100 simulations. The results for the developed method are presented in Fig. 6. Since the developed method uses a very simple baseline, it makes sense to compare it with a more advanced baseline model. The model used for comparison is a baseline function based on linear interpolation with exponential smoothing. The results corresponding to Fig. 6 are in Fig. 7. Averaging the RMSE for 100 iterations seems to give relatively smooth graphs, while a smaller number of iterations could cause single outliers to have a large effect on averages and the results.

The noise in the data affects the goodness of the results somewhat linearly. This is explained by the baselines being increasingly weak in predicting volumes as the noise increases. The errors could possibly be decreased using a longer time window. A better baseline directly decreases the errors in the measured volume changes, while a longer time window would allow more data points, resulting in increased reliability for the model.

Another observation is that a higher number of products increases the error in the results with both baseline methods. This results from the difficulty of allocating the cannibalization effect to the correct products with noisy

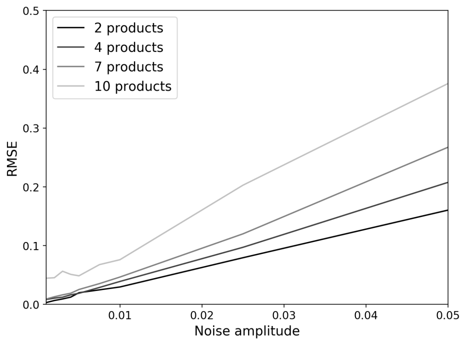

Figure 6: Results of the simulated runs.

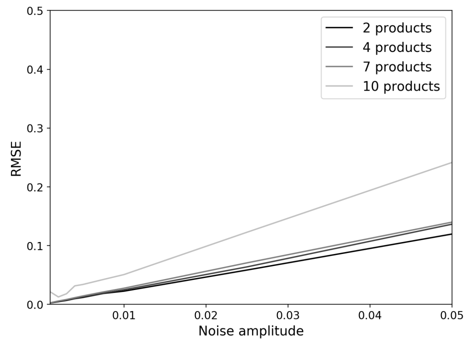

Figure 7: Results of the simulated runs with better baselines.

data. Without noise, it is possible to find an exact solution in the simulated data. However, as soon as noise is introduced, the volume changes become somewhat unreliable. If the uplift for a promoted product is clearly higher than what it would be in an ideal data for a single individual promotion, it is calculated to heavily cannibalize other products. This lowers the estimated cannibalization for other promoted products. These errors then cause similar effects in other promotions and so on. Because of the regularization, this effect is mitigated as large weights are penalized, but this nevertheless illustrates why the number of products increases errors.

Comparing Fig. 6 and Fig. 7, it is clear that a better baseline model significantly decreases the errors in the results. This results mainly from the volume changes better modeling the real change caused by promotion when comparing to a baseline that takes the trend and seasonality into account. Another improvement is that unlike the naive baseline model, a model with smoothing makes it possible to utilize all promotion weeks in the calculation, as opposed to only the first week of each promotion, thus increasing the available data points with the same time window. As calculating the uplifts and downlifts is done in a separate function, changing between different methods is straightforward, as long as the required fields (volume baseline in this case) are evaluated first.

Shortening the time window in half to 1.5 years decreases the accuracy significantly, as seen in Fig. 8 and Fig. 9. This simulation uses only 50 iterations per data point in order to reduce the running times, but the results should still be representative, even if the curves are slightly less smooth. In Fig. 8 it is visible that for 10 products, the method fails even on low levels of noise. This is likely to be a result of too little data to determine which products are the actual cannibals. For this data, the results with the proper baseline

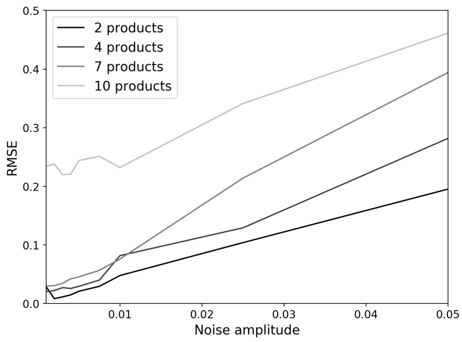

Figure 8: Results of the shorter simulated runs.

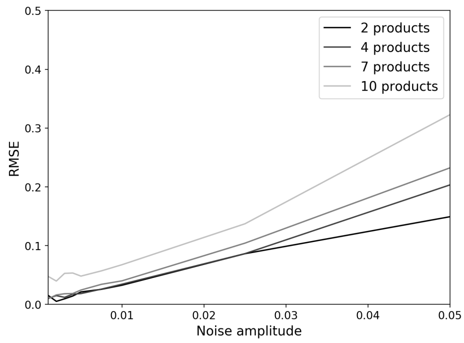

Figure 9: Results of the shorter simulated runs with better baselines.

model fall to the same level as for three years with simple baselines. This is a direct consequence of having fewer data points to determine the cannibalization from. This way, it becomes increasingly difficult to smoothen out the noise from the demand changes, resulting in a serious decrease in the reliability of the method. For the simple baseline, the situation is even worse, and the spread of the cannibalization estimates over 50 runs is plotted in a box plot in Fig. 10. This way it can be seen that the 50% confidence intervals for the coefficients are wide, implying that the results for a single run are unreliable. The median values are relatively close to the correct value shown on the x-axis, but for the largest cannibalizations (-0.25), even the median is relatively far from the correct value.

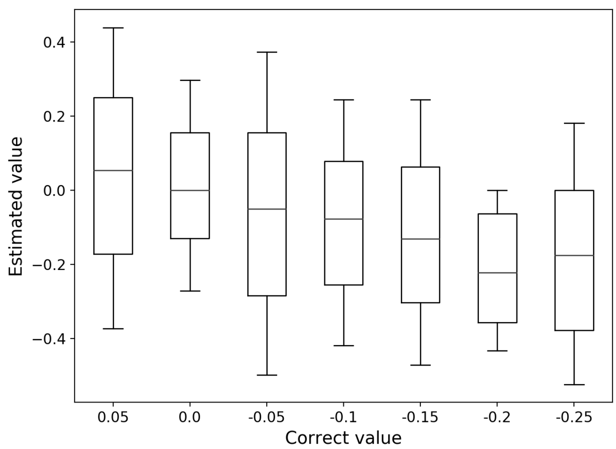

Figure 10: A box plot of the coefficients for 7 products and a noise level of 0.05.

The RMSE as a function of noise in Figs. 6-9 seems approximately linear, except for one in Fig. 8, and the method is accurate without noise (unless there is no promotion data for a product, which results in the cannibalization estimate always being zero). Therefore, we fit a linear regression $y = kx$ for all the graphs, and plot the slopes k in Fig. 11. This way, we can visually confirm the performance ranking of the different scenarios. It can be seen that the baseline method with linear interpolation and smoothing is roughly 40%-50% better than the simple naive method, and 3 years of data 30%-40% better than 1.5 years.

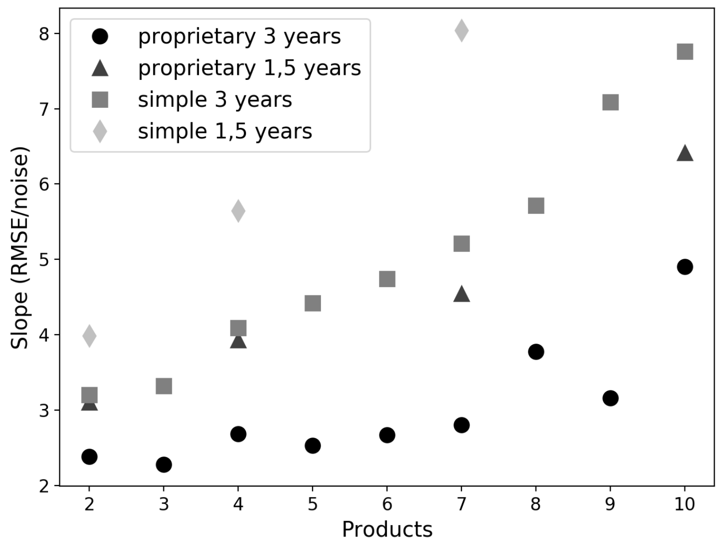

Figure 11: Comparison of the baseline methods.

The main problem with testing on a simulated dataset is that both the data and the method have the same underlying assumptions. What this means is that we can only say how the method performs assuming that our assumptions are correct. The performance on actual retailer data is likely to be worse than the results imply, as consumer choice in reality will not follow the assumptions as strongly as in the simulated data. Even with this problem, the results are still certainly useful, as they show which item pairs have a strong cannibalization or complementarity between the items. The magnitude of the effect should be taken as an estimate, but considering the scarcity of results with a similar approach in the literature, this work does contribute to the research on cannibalization between products.

5 Conclusions

The main object of this thesis was to develop a method for evaluating the magnitudes of cannibalization for item pairs from sales data. The main goals for the method were accuracy and stability with reasonable running times. Of these goals, stability was most clearly achieved, as the algorithm converges to the same solution independent of the initial guess.

The accuracy of the method was examined in the previous section, and the main finding was that while the errors do not increase extremely fast as the noise in the data increases, the simple baseline model performed considerably weakly. However, the algorithm is built in such a way that the baseline model is easily changed if better ones are developed.

Running times were relatively good, and further speedup was not attempted in the context of this thesis. However, converting the algorithm from Python to C or Scala and parallelizing the processes, vast improvements to running times could be achieved in multiple parts of the algorithm. These improvements might be necessary if the number of products examined increases further, as the number of coefficients estimated is N^2. Another option would be to use a more powerful computer, but parallelization gives true scalability as Eq. 3 could be solved in N separate processes and the results for each store could be calculated independently. Additionally, the regularization hyperparameters λ_1 and λ_2 could be fine-tuned to achieve optimal convergence rates without oscillation.

According to The Boston Consulting Group, the overall effect of promotion planning could be an increase of 2 to 5 percent points in promotion margin (Goad et al. [2015]). As a rough example, a company with 100M€ revenue of which 20% comes from promotions, would gain at least $100M€ \times 20\% \times 2\% = 0,4M€$ annually. Cannibalization is seen as a large driver in promotion effectiveness, and it is safe to say that the potential impact of understanding how different products cannibalize is significant for a large retailer.

For retailers, the average sales cannibalization for a product or a category has a significant meaning in properly understanding their promotions. However, due to the behavior described in Fig. 2, converting volume cannibalization to sales cannibalization is not trivial in promotion cannibalization. The approach based on unit price ratios is valid as long as the unit prices stay constant. However, in our case that requirement is not necessarily fulfilled. Thus, the best way to get category averages would be to calculate the sales cannibalization separately for each item and promotion. This would be done by first using volume cannibalization, then calculating the total sales uplift and cannibalization drop. From these results, calculating an estimate of the overall category cannibalization with the current marketing mix would be trivial.

5.1 Future work

Although the method performs relatively well on simulated data, there are clear problems when we take a look at actual sales data. In the simulated data, cannibalization always behaved according to the assumptions. However, even if the assumptions work for regular price-cut promotions, retailers do a wide range of different promotions. The first problem is so called multibuys, where a discount is applied to a set of products. When a multibuy offer contains different items (e.g. a selection of frozen pizzas of a certain brand), there is no cannibalization between them, even though they have a large potential for cannibalization in regular promotions due to their similarity. These promotions could be dropped from the data, but that approach also has its disadvantages, as the promotion could be very significant. Simply excluding the multibuy products would result in their cannibalization being allocated to other simultaneously promoted products. Another alternative would be to exclude the whole week from the data, but this would quickly lead to the data becoming smaller, resulting in reduced reliability.

Another challenge arising from different promotion types is that, as mentioned earlier, a TV ad causes consumer behavior different from an in-store promotion. What should be examined is whether this difference in cannibalization is a constant multiplication across all products. If this is the case, a deeper understanding of the phenomenon would be achieved.

In order to avoid excessive calculations, the dataset must be pruned before using the method. For example, the cannibalization between ice cream and carrots is probably insignificant. For this pruning, a clustering algorithm, such as the well-known k-means, could be used to group items with similar properties. It must be noted however that excluding significant items from the dataset is worse than including insignificant items. If a product causing major cannibalization is excluded, the cannibalization is allocated to other cannibalizing products, resulting in errors, while including irrelevant products should only increase running times.

References

Andrew Abere, Oral Capps, Jeffrey Church, and Alan Love. Mergers and Market Power: Estimating the Effect on Market Power of the Proposed Acquisition by the Coca-Cola Company of Cadbury Schweppes' Carbonated Soft Drinks in Canada. In *Measuring Market Power*. 2002.

Rakesh Agrawal, Tomasz Imielinski, Arun Swami, Harry Road, and San Jose. Mining Association Rules between Sets of Items in Large Databases. page 10, 1993.

Robert C. Blattberg and Kenneth J. Wisniewski. Price-Induced Patterns of Competition. *Marketing Science*, 8(4):291–309, 1989. ISSN 0732-2399. URL http://www.jstor.org/stable/183977.

Lee G. Cooper. Competitive Maps: The Structure Underlying Asymmetric Cross Elasticities. *Management Science*, 34(6):707–723, 1988. ISSN 0025-1909. URL https://www.jstor.org/stable/2632125.

dunnhumby. PriceStrat, 2015. URL https://www.dunnhumby.com/sites/default/files/filepicker/1/dunnhumby_PriceAndPromotions_Promotions_Optimisation_Brochure.pdf.

Robert H Frank. *Microeconomics and Behavior*. McGraw-Hill/Irwin, 2008. ISBN 978-0-07-126349-8.

Nick Goad, Jeff Robinson, Javier Anta Callersten, Andreas Malby, and Jacob Opstrup. How Retailers Can Improve Promotion Effectiveness, July 2015. URL https://www.bcg.com/publications/2015/retail-pricing-how-retailers-can-improve-promotion-effectiveness.aspx.

Harald J. Van Heerde, Peter S. H. Leeflang, and D. R. Wittink. *Flexible Decomposition of Price Promotion Effects Using Store-Level Scanner Data*. 2002.

James L. Heskett. *Marketing*. Macmillan, 1976. ISBN 978-0-02-353940-4. Google-Books-ID: sO0TAQAAMAAJ.

John DC Little. Brandaid: an on-line marketing-mix model. 1972.

Wendy Lomax. The measurement of cannibalization. *Marketing Intelligence & Planning*, 14(7):20–28, December 1996. ISSN 0263-4503. doi: 10.1108/02634509610152673. URL https://www.emeraldinsight.com/doi/10.1108/02634509610152673.

Charlotte H. Mason and George R. Milne. An approach for identifying cannibalization within product line extensions and multi-brand strategies. *Journal of Business Research*, 31(2-3):163–170, October 1994. ISSN 01482963. doi: 10.1016/0148-2963(94)90080-9. URL http://linkinghub.elsevier.com/retrieve/pii/0148296394900809.

Xia Ning and George Karypis. SLIM: Sparse Linear Methods for Top-N Recommender Systems. page 10, 2011.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Sundara Raghavan Srinivasan, Sreeram Ramakrishnan, and Scott E. Grasman. Identifying the effects of cannibalization on the product portfolio. *Marketing Intelligence & Planning*, 23(4):359–371, June 2005. ISSN 0263-4503. doi: 10.1108/02634500510603465. URL https://www.emeraldinsight.com/doi/10.1108/02634500510603465.

RELEX. Cannibalization and Halo Effects in Demand Forecasts, January 2018. URL https://www.relexsolutions.com/cannibalization-halo-effects-in-demand-forecasts/.

Revionics®. Revionics® Promotion Optimization. URL http://pages.revionics.com/rs/343-ELA-089/images/DataSheet_PromoOpt_v2.pdf.

Shuba Srinivasan, Koen Pauwels, Dominique M. Hanssens, and Marnik G. Dekimpe. Do Promotions Benefit Manufacturers, Retailers, or Both? *Management Science*, 50(5):617–629, 2004. ISSN 0025-1909. URL https://www.jstor.org/stable/30046102.

Sean J Taylor and Benjamin Letham. Forecasting at scale. page 25, September 2017. doi: 10.7287/peerj.preprints.3190v2. URL https://peerj.com/preprints/3190.

Rockney G. Walters. Assessing the Impact of Retail Price Promotions on Product Substitution, Complementary Purchase, and Interstore Sales Displacement. *Journal of Marketing; Chicago*, 55(2):17, April 1991. ISSN 00222429. URL https://search.proquest.com/docview/227765721/abstract/53D26CAC8AAA4EFCPQ/1.

Yan Yuan, Oral Capps, and Rodolfo M. Nayga. Assessing the Demand for a Functional Food Product: Is There Cannibalization in the Orange Juice Category? *Agricultural and Resource Economics Review*, 38 (02):153–165, October 2009. ISSN 1068-2805, 2372-2614. doi: 10.1017/S1068280500003178. URL https://www.cambridge.org/core/product/identifier/S1068280500003178/type/journal_article.