# Clustering Stores of Retailers via Consumer Behavior

Thesis in Business Analytics and Quantitative Marketing

*Author:*
Gijs van Rooij (360497)
*Supervisor:*
Gertjan van den Burg
*Co-reader:*
Dr. Michel van de Velden

Econometrics & Operations Research
Erasmus School of Economics
Erasmus University Rotterdam

February 2, 2017

# Abstract

These days most retailers define clusters of stores and set different prices in the clusters, since it is not yet attractive to define store-by-store prices due to optimization and operational issues. However, retailers define clusters of stores solely based on local competition. Existing clusters of stores could be further broken down and price management decisions could be adjusted accordingly by using consumer behavior. Therefore, the price elasticity is used which is a reflection of consumer behavior. In this way the retailer enables itself to set different prices in smaller clusters of stores in order to attain higher revenue and profit. This research provides a clustering solution that defines clusters of stores based on price elasticities. This clustering solution proves potential with a projected increase in revenue of 0.36% and a projected increase in profit of 0.76% by using data of one of the major supermarket chains in the Netherlands.


**Keywords:** *clustering, constrained clustering, cluster consensus, NACT*

# Contents

# 1 | Introduction

In this chapter we propose the problem at hand and we briefly introduce the methods applied to answer the research question. First, in Section 1.1 the topic of the research alongside the research question and the context of the research are provided. Consequently, in Section 1.2 the workflow of the research is described. Then, the applied methods are briefly discussed in Section 1.3 and in Section 1.4 the structure of the paper is discussed.

## 1.1 | Topic and data description

Since it is not yet attractive – for optimization and operationally for managers – to define store-by-store prices, retailers assign stores to clusters of stores. Retailers create clusters of stores to set different prices for the same products in different clusters. Stores assigned to the same group charge the same prices for the same products. Unfortunately many suppliers rely too heavily on their internal perceptions with regard to their clustering policy [29]. These days most retailers define clusters of stores solely based on local competition whereas the variation in consumer behavior among stores is completely neglected. Obviously, consumer behavior varies between different stores. Variance in consumer behavior leads to differences in demand and therefore it is attractive for retailers to set different prices in different stores for the same product. The aim of this research is clustering stores based on price elasticity, which is a reflection of consumer behavior, to prove that there is an opportunity to increase revenue and profit by using a new way of clustering. In conclusion, since retailers in general exclusively focus on local competition a great opportunity gets lost. This leads us to address the following research question:

**How can we cluster stores by means of the differences in consumer behavior such that the revenue and profit of the retailer increases?**

The context of this research is supermarkets. We define a clustering framework using data of the Jumbo supermarket chain. Jumbo owns over 600 stores spread all over the country which makes them one of the major players in the Netherlands with regard to grocery shopping. According to Nielsen marketing the market share amounted approximately 20% in 2015[1]. Jumbo defines four different clusters of stores based on local competition. However, there is still a lot of variation in price elasticity within the clusters of stores and especially in some specific groups of products (PGs) between stores. For example, the PG Cola consists of all products related to Cola such as bottles of different volume from Pepsi and Coca Cola. Each product group is defined by the retailer. This provides us with an opportunity to boost sales and profit by defining smaller subsets of 'similar' stores within existing clusters of stores based on the price elasticities.

[1]http://www.distrifood.nl/service/marktaandelen

## 1.2 | Workflow

The workflow of this research consists of three phases, namely finding clusters of stores for each specific product group (phase 1), obtaining the final clusters of stores (phase 2), and analyzing for the impact of the obtained clusters of stores (phase 3). Each of these phases raise different questions that need to be answered to get to the answer of the research question.

### 1.2.1 Phase 1: Clusters of stores per product group

First we obtain clusters of stores for each product group separately, since consumer behavior among product groups differs [30]. Multiple clusterings impose different structures on the data and therefore provide a wide range of information. As mentioned before, the price elasticity reflects the consumer behavior. We use the price elasticities per store per product which are called the price elasticities on store-item level. Since the price elasticity on store-item level is the most specific level of price elasticities available, no information gets lost in the process of aggregating to a higher level of price elasticities such as on store level. In this case the information per product is no longer considered and therefore the information is less comprehensive. Each store is assigned to a seperate cluster of stores based on the store-item level price elasticities of all available products. In order to answer the research question we first need to answer the following subquestion:

**How can we cluster the stores per product group exclusively based on the products available in the considered product group?**

### 1.2.2 Phase 2: Final clusters of stores

Obviously, in phase 1 stores could end up in a different cluster of stores for the different product groups. For example, store $s$ could end up in the first cluster for product group $g$ and in the second cluster for product group $h$ even after relabeling. In the end the retailer searches for one specific set of clusters of stores instead of a clustering per product group due to operational and optimization limitations. We could think about problems such as the replacement of paper price tags and the computing speed limitations of computers whereby price optimization could take ages. Therefore we need to combine the information comprised in the clustering per product group into one final set of clusters of stores. This leads us to address the following subquestion:

**How can we obtain one final set of clusters of stores that combines the information comprised in the clusters of stores found for each separate product group?**

### 1.2.3 Phase 3: Clustering impact

Price is the factor that provides immediate revenues [27]. Therefore, to prove the positive impact of our framework on the revenue and the profit we have to adjust the current prices according to the new formed clusters. We have to define a price policy at which we set different prices for products in different clusters of stores. The price of a product in a cluster is kept the same in all stores in that cluster. Note, the focus of this research is not to find the optimal prices but finding the clusters of stores that fit the underlying data the best. This raises the following subquestions to answer the research question:

**If different prices are set in the clusters, what is the impact of the clustering of stores for each of the separate product groups on the revenue and profit of these product groups?**

**If different prices are set in the clusters, what is the impact of the final clustering of stores on the revenue and profit in total?**

## 1.3 | Method

The main focus of this research is to define new clusters of stores that use the differences in consumer behavior among stores. To use differences in consumer behavior by means of the variance in price elasticity on store-item level a two-stage framework is needed. First, in phase 1 and the first step of the two-stage framework we obtain clusters of stores based on the price elasticity on store-item level for each product group separately. The K-means++ clustering algorithm by Arthur and Vassilvitskii [2] is used as a building block for the Restricted K-means++ clustering algorithm which is proposed in this paper. The K-means++ clustering algorithm is an easily implementable clustering algorithm that is able to handle large datasets. The proposed Restricted K-means++ clustering algorithm is bound to some restrictions for which K-means++ does not comply and therefore the Restricted K-means++ clustering algorithm is an extension of the K-means++ clustering algorithm. Stores should be assigned to clusters such that stores originating from the same city/village are appointed to the same cluster. If stores from the same city/village end up in different clusters our reassignment process ensures these stores in the end are assigned to the same cluster. Furthermore, products that generate more revenue than others are more important to the retailer. This is covered in the novel weighted dissimilarity measure between stores proposed in this paper. The higher the weight, the more important the product is to the retailer. Moreover, imagine we are interested in the dissimilarity between two stores and only one of the stores sells a specific product. Although the price elasticity of the product is not available in one of the stores it indicates a difference between the stores because the assortment differs. To cope with this issue we add a correction term to the novel weighted dissimilarity proposed in this paper that scales up the dissimilarity in case data is missing. Secondly, phase 2 and the second step of the two-stage framework includes combining the information comprised in the clusterings of stores from the separate product groups by using the hard least squares Euclidean consensus algorithm [14]. The result is the so-called cluster consensus which represents the final set of clusters of stores. Finally, we conduct an extensive analysis using regressions which provides insights in the drivers for differences in consumer behavior in order to make better decisions.

## 1.4 | Structure paper

In the remaining of this paper, first the related literature is discussed in Chapter 2. In Chapter 3 the data used for evaluation is analyzed and in Chapter 4 the proposed method regarding clustering is described. The results of the clustering are provided in Chapter 5. In Chapter 6 the conclusion and some suggestions for future work are presented.

# 2 | Related work

In order to use the variation in price elasticities we propose a two-stage clustering framework which consists of first finding new clusters of stores for each product group separately and consequently finding the cluster consensus. The cluster consensus summarizes the information of all the found clusters of stores for each product group separately in one final set of clusters. Additionally, we conduct an extensive analysis using regressions on the formed clusters which provide insights in the drivers for differences in consumer behavior in order to make better decisions. Existing methods are re-used and new ones are introduced. Section 2.1 presents an overview of the literature on clustering. In Section 2.2 we elaborate upon the literature related to the drivers of price elasticity.

## 2.1 | Clustering

The K-means clustering algorithm [12] is a clustering method that is the most commonly used clustering algorithm in particular for its simplicity and its applicability to large data sets. K-means suffers from the limitation that convergence to a local optimum could produce unreliable results due to the choice of the initial centers. As such, the K-means++ algorithm was proposed by Arthur and Vassilvitskii [2] to avoid the problem of getting stuck in a local optimum. Moreover, the K-means++ algorithm outperforms the K-means algorithm regarding speed as well. Additionally, in this research we are bound to the domain-specific business rule that stores located in the same city/village should be assigned to the same cluster. This business rules can be implemented in the clustering algorithm via constraints. Related to it, constrained clustering [3, 4, 5] has been widely researched the last decades.

Wagstaff et al. [40] implement constraints into the K-means clustering algorithm by ensuring none of the constraints are violated while updating cluster assignments. In case point $s$ cannot be hosted by cluster $C$, they go through the list of clusters and search for a cluster to which point $s$ could be assigned. The must-link constraint, i.e., two instances should be in the same cluster, corresponds to the business rule our research is bound to. The disadvantage is that the algorithm is order-sensitive since the order of cluster centers to which the stores are compared could influence the final clustering. Therefore the final clusters could end up differently in separate runs if the order of cluster centers differs.

Basu et al. [3] needs supervisory information given as cluster labels a priori to initialize the clustering algorithm. They propose to incorporate this form of semi-supervision in the K-means algorithm by seeding. It uses labeled data and the constraints generated from labeled data to initialize the clustering algorithm. This way so-called seed clusters are generated to initialize the clustering algorithm. The labels of the seeded data are kept the same during subsequent steps of the algorithm, whereas the non-seeded data labels are estimated at each step. Additionally, they propose to apply the same technique using the Expectation Maximization algorithm. During each Maximization step the conditional distributions are kept the same for the seeded data and at each Expectation step the conditional distributions for the non-seeded data are re-estimated. Experimental results on a newsgroups dataset show the convenience of the aforementioned algorithms regarding sensitivity and robustness over random seeding and COP-Kmeans, proposed by Wagstaff et al. [40].

Another method for constrained clustering is the Constrained Complete Link algorithm by Klein et al. [24]. This hierarchical agglomerative clustering method using complete linkage is altered during initialization to cope for the must-link and cannot-link constraints, i.e., two instances should not end up in the same cluster. In case instances are related by a must-link constraint the distance between the instances is set to zero, while in case of the cannot-link constraint the same distance is set to infinity. Consequently, new distances between pairs of instances are found using shortest path.

In contrast to constructing a single clustering, multiple clusterings impose different structures on the data and therefore provide a wide range of information. In order to exploit the complementary nature of the data, different samples of input data can be used. To combine clusterings, according to Vega-Pons and Ruiz-Scholcloper [39], roughly two sorts of different methods exist, i.e., median partition based approaches [13, 39] and object co-occurrence based approaches [14, 22, 38]. The basic idea of a median partition based approach is to find a clustering $P$ such that the similarity between $P$ and all the clusterings in the ensemble is maximized. One of the object co-occurrence based approaches is the method by Strehl and Ghosh [36] for which first the similarity matrix is constructed. Based on the adjacency matrix of a hypergraph the similarity matrix can be deduced. If the adjacency matrix contains a 1 in the column it represents an object (row) that is contained in that cluster. Next, any reasonable similarity-based clustering algorithm can be used, such as K-medoids [23]. However, since the adjacency matrix of a hypergraph is constructed they also propose to use the METIS algorithm [21] to yield the combined clustering which then makes the algorithm a graph based method.

## 2.2 | Covariates of price sensitivity

The drivers of consumer behavior are especially interesting to the retailer to explain differences in price elasticities between products and stores. Shifts in the demand curve are caused by main determinants such as the disposable income of households and prices of related goods. Along with these determinants, it is expected that the explanatory power of the price elasticity rises by means of many other components. To define the covariates of price elasticity a wide range of literature can be found.

Hoch et al. [16] managed, in contrast to the unsuccessful efforts of many other researches [1, 34], to find a relationship between consumer characteristics and price elasticity using scanner data of 18 product groups in 83 stores in Chicago. They discovered they could explain two thirds of the variation in price elasticity by using eleven consumer characteristics consisting of demographic and competitive variables. The information comprised in seven demographic variables related to life stage, education, family size, income, ethnicity, value of the house, and the percentage of working women are exploited. Moreover, four competitive variables such as the sales of competitors and distances to other supermarkets and warehouses are incorporated in the model. The demographic variables proved to be more important than the competitive variables. However, the research dates from 20 years ago and the retailing market has changed drastically. Collecting data is way less expensive these days and therefore the data could contain more valuable information. Moreover, since the data was collected in Chicago the competitive landscape and consumer behavior are expected to differ from the competitive landscape and consumer behavior in the Netherlands.

As stated before, price elasticity is broadly researched and all sorts of different determinants to price elasticity are examined. Kumar and Karande [26] seek the effect of the environment of a store on the retailer's performance by means of the sales or productivity of stores located all over the United States. Most interestingly, they conducted the analysis on store level and segmented stores offering useful insights in the determinants of store performance. Situational price sensitivity is examined by Wakefield et al. [41]. They found in case of social and hedonic consumption situations price sensitivity is weakened compared to functional consumption. Huang et al. [19] found that low-income shoppers appear to be most price sensitive and price sensitivity decreased in case market share of brands increased. Last, Elrod and Winer [11] found weak significant influence of age, education, and female headed households on price sensitivity.

# 3 | Data

As we know retailers assign stores to clusters. Stores assigned to the same cluster charge the same prices for the same products. As previously mentioned in Chapter 1 Jumbo implements four unique clusters of stores based on local competition. Each of the formed clusters of stores is called a priceline. The stores are assigned to a priceline based on the nearest located competitor in the same domicile. In this research we focus on the priceline represented by the largest competitor of Jumbo which is Albert Heijn due to availability issues of the data. Within this priceline we conduct a clustering of stores. This cluster of stores contains 305 stores out of the over 600 stores in total. These stores account for 53% of the revenue of the Jumbo market chain. Stores that opened at most 3 months ago are discarded from the research due to the limited amount of historical data. Furthermore, only products which are sold in the last quarter (2016Q1) of the considered time span are taken into account since products which are not sold anymore do no longer influence the policy of a retailer. There are two sources of data used: the Jumbo database and Statistics Netherlands (CBS). We use CBS data to retrieve information on demographics related to the environment of the considered stores. The price elasticities on store-item as well as on store level are provided by the retailer. Moreover, the data is roughly split into two sets, the scanner data analyzed in Section 3.1 and the store-specific characteristics and trading area data discussed in Section 3.2.

## 3.1 | Scanner data

We retrieve weekly scanner data at store-item level for a time span of 156 weeks from the Jumbo database. We focus only on the non-alcoholic beverages again due to availability issues. Furthermore, in general the product groups related to non-alcoholic beverages show most variation in price elasticity. Therefore these product groups are most promising regarding a boost in sales and profit. The non-alcoholic beverages account for 5% of the euro sales volume. The considered product groups are displayed in Table 3.1. The average price elasticity per product group per store is found by dividing the price elasticity per product group per store by the number of stores that sell products of the considered product group. The price elasticity per store per product group is provided by the retailer. It can be easily seen that the product group Juices has the highest average price elasticity per store, whereas Cola has the lowest average price elasticity per store. On average the product groups contain 158 different products with 239 at a maximum for the product group Kids youth and 66 at a minimum for the product group Energy.

**Table 3.1:** The evaluated product groups with the number of products, number of brands, average price elasticity per store and standard deviation of the price elasticity.

| Product group | Number of products | Number of brands | Average elasticity | Standard deviation |
|---|---|---|---|---|
| Cola | 108 | 6 | -0.946 | 0.180 |
| Energy | 66 | 18 | -0.964 | 0.116 |
| Juices | 233 | 28 | -1.509 | 0.167 |
| Kids youth | 239 | 32 | -1.056 | 0.130 |
| Large soda | 204 | 25 | -1.044 | 0.139 |
| Water | 95 | 14 | -1.016 | 0.096 |

This scanner data set comprises the retail prices, prices for substitutes and complements, as well as information on promotions such as price discounts. The extensive data on promotional activity also incorporates indicators for the promotional channels such as local media, flyers, and television commercials. Furthermore, the data contains information on holidays, weather, and school vacations.

## 3.2 | Store-specific characteristics and trading area data

The Jumbo database and CBS provided us with data on store specifics and trading area data which are composed of competition, socio-demographics, and socio-economic variables. All stores as accounted for in the scanner data set are represented. Moreover, to get a grasp into the reasons for variation in price elasticity we retrieve information from the database on store-specific characteristics such as franchise indicators, store offerings being for example information on the presence of self scanners, the surface in square meters of a store, the city-rural dummy and many more. The city-rural dummy is useful to cope for population density within the trading area of specific stores. The trading area is determined by competition among others. The competition is split into two separate groups: supermarkets and miscellaneous such as liquor and drug stores. The retailer defines the competitive environment of stores by collecting information on the competitive supermarkets and the miscellaneous competitors within a radius of 15km from each store. Note, the stores within a radius of 15km that belong to the supermarket chain of Jumbo itself are also indicated as competitors since sales of these specific stores could transfer to another Jumbo store due to the presence of another 'competitive' Jumbo store. The considered socio-economic variables comprise variables such as education, income, and social class. The trade area characteristics take also socio-demographic variables into account like household size. Additionally, from CBS we retrieve information on ethnicity in the form of the percentage of western and non-western immigrants.

# 4 | Method

This research strives to use the variance in price elasticity to get to new clusters of stores by conducting a two-stage clustering framework. Figure 4.1 shows there is still a lot of variation in price elasticity within clusters of stores and especially in some specific product groups between stores which proves the need for a new method.

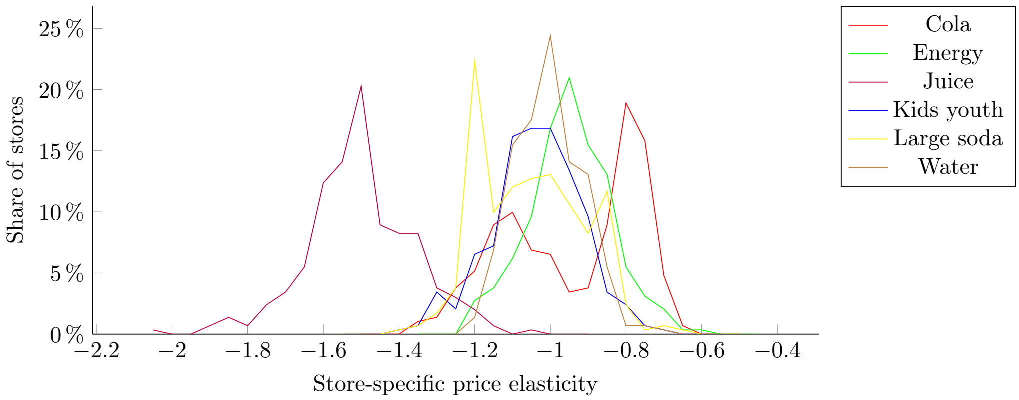


**Figure 4.1:** The share of stores by the store-specific price elasticity.

Differences in price elasticities on store-item level are studied [32] to find new clusters of stores. We propose a novel dissimilarity measure in Section 4.1 that is used in the clustering framework. Next, we discuss the clustering framework and the evaluation and validation methods in Section 4.2. Finally, Section 4.3 shows an outline of the regression model that is used for the analysis of the final set of clusters found by the clustering framework. Note, we assume that each product only belongs to one product group. In the remainder of this paper symbols $i$ and $j$ refer to products, whereas $s$, $v$, and $y$ indicate stores. The set of all stores is indicated by $S$ and the set of all products by $I$. Furthermore, $G$ is the set of product groups. Then,

$$I = \bigcup_{g=1}^{G} P_g \tag{4.1}$$

where $P_g$ is the set of products in the product group with index $g$. The input data for the clustering algorithm $\Gamma = \{e_{i,s} \ \forall i \in I, s \in S\}$ comprises all price elasticities $e_{i,s}$ per product $i$ in store $s$. The clustering function for each product group indicated by $g$ can be described as:

$$\lambda_g : S \rightarrow \{1, ..., k_g\}, \ \forall g \in \{1, ..., G\}, \tag{4.2}$$

with $k_g$ is the number of clusters for product group $g$ and subsequently the set of clusterings is defined as follows:

$$\Lambda = \{\lambda_g : g = 1, ..., G\}. \tag{4.3}$$

This way we define a clustering of stores for each product group separately via $\Lambda$ based on the price elasticities on store-item level. The information comprised in $\Lambda$ is used to find the final optimal clustering $\lambda^*$ by means of the hard least squares Euclidean consensus algorithm. In the appendix, Table A.1 shows an overview of all symbols used for the dissimilarity measure and the clustering method.

## 4.1 | Dissimilarity measure

For the dissimilarity measure between stores, which is used in the clustering method, two conditions should be taken into account: first *product importance* and second *not available values* (NA). That is why the commonly used Euclidean distance for clustering does not suffice as dissimilarity measure. First obviously, products with relatively high revenue are most valuable to the retailer which makes us introduce weighted dissimilarities. The higher the weight, the more important we consider a specific product. Weight $w_i$ is defined as follows:

$$w_i = \frac{\overline{\psi_i} d_i}{\sum_{i \in N(s,y)} \overline{\psi_i} d_i}, \tag{4.4}$$

where $N(s, y)$ is the set of all products in a product group sold in both stores $s$ and $y$, $\overline{\psi_i}$ equals the average revenue per store for product $i$, and $d_i$ represents the number of stores in which product $i$ is sold. Weight $w_i$ corresponds to the total revenue of product $i$ divided by the total revenue of all products sold in both stores $s$ and $y$ in its corresponding product group such that $0 \leq w_i \leq 1$. Consequently, the result is the following preliminary weighted dissimilarity measure between store $s$ and store $y$:

$$D'(s, y) = \sum_{i \in N(s,y)} w_i |e_{i,s} - e_{i,y}|, \tag{4.5}$$

where $|e_{i,s} - e_{i,y}|$ is the absolute difference between the price elasticities of product $i$ sold in store $s$ and store $y$. The sum is exclusively evaluated over the set $N(s, y)$ since it is only possible to subtract price elasticities in case both stores sell the product. If a product is solely sold in one of both stores, the price elasticity of the product is not available in one of the stores and we can not subtract the price elasticities of the products as simply one of them misses. Note, if both stores do not sell the considered product there will be no difference in assortment and therefore there is no indication for a difference between the stores regarding the considered product.

Second, again imagine we are interested in the dissimilarity between two stores and only one of the stores sells a specific product. Although the price elasticity of the product is not available in one of the stores it indicates a difference between the stores because the assortment differs. Therefore, Porro et al. [33] suggest to address the problem before the classifier is built. They propose to adjust the dissimilarity measure to cope for the not available values by using the available data only. The Not Available Correction Term (NACT) is proposed to correct for products which are solely sold in one of the two considered stores. The NACT is represented by the fraction $\psi_{\cup}/\psi_{\cap}(s, y)$ and consists of the following terms:

$$\psi_{\cup} = \sum_{j \in P_g} \overline{\psi_j} d_j \tag{4.6}$$

and

$$\psi_{\cap}(s, y) = \sum_{i \in N(s,y)} \overline{\psi_i} d_i, \tag{4.7}$$

where $\psi_{\cup}$ is the total revenue of all products sold in a product group, whereas $\psi_{\cap}(s, y)$ is the total revenue of all products exclusively sold in both stores. Accordingly, the final domain-specific dissimilarity measure is as follows:

$$D(s, y) = \frac{\psi_{\cup}}{\psi_{\cap}(s, y)} \sum_{i \in N(s,y)} w_i |e_{i,s} - e_{i,y}|. \tag{4.8}$$

This way the NACT is responsible for the relative importance of the set of products sold in both stores in a product group versus the set of products exclusively sold in one of both considered stores in a product group. The following hypothetical example provides insight in the NACT. Imagine we compare two stores $s$ and $y$. There are three products sold in the union of products sold in stores $s$ and $y$ in the considered product group. Product 1 is exclusively sold in store $s$ and has a revenue of 10 in total. Moreover, products 2 and 3 have a revenue of 20 and 30 in total in both stores, respectively. Therefore, the numerator of the NACT equals 60 and the denominator equals 50 which results in a NACT of $60/50$. This means the dissimilarity measure is scaled up since one of the products is exclusively sold in one store. We want to take this products as well into account with regard to the dissimilarity between two stores since the assortment differs. Finally, the weight $w_i$ addresses the relative importance of products sold in both stores relative to each other. Note, the denominator of the NACT and weight $w_i$ are the same, i.e, $\psi_{\cap}(s, y)$.

## 4.2 | Clustering

This section provides the two-stage clustering framework along with the method for clustering validation and evaluation. Subsection 4.2.1 introduces the novel Restricted K-means++ clustering algorithm and Subsection 4.2.2 the process of finding the cluster consensus. Subsection 4.2.3 enumerates the methods for determining $k$ the number of clusters. The cluster evaluation method with regard to robustness of the algorithm and the price policies are shown in Subsection 4.2.4.

### 4.2.1 Restricted K-means++

*K-means++ algorithm*

K-means aims to partition $n$ observations into $k$ clusters such that each observation is assigned to the closest cluster center that serves as a prototype for the cluster. As K-means could produce unreliable results due to the choice of the initial centers, K-means++ proposes to use the randomized seeding technique that copes with the issue of ending up in local optima. This way initial centers are seeded at random. At first initial center $c_1$ is picked uniformly at random concerning the randomized seeding technique. Next, let $C_l = \{c_1, ..., c_l\} \subseteq C$ be the current set of cluster centers. Then, store $y$ is the next cluster center $c_{l+1}$ with probability:

$$P(c_{l+1} = y) = \frac{\min\limits_{c \in C_l} D(y,c)^2}{\sum_{s \in S} \min\limits_{c \in C_l} D(s,c)^2}, \; \forall y \notin C_l, \tag{4.9}$$

where $S$ is the set of all stores, $C_l$ is the current set of cluster centers, and $\min_{c \in C_l} D(y,c)$ is the shortest distance of store $y$ to the closest center $c$ that is already chosen. This way if the distance of a store to the closest center $c$ that is already chosen becomes higher, the higher the probability becomes that this store is chosen as the following cluster center $c_{l+1}$.

*Cluster centers*

As stated before we have to deal with products which are not sold in certain stores resulting in values which are not present in the data. Apart from the problem these values cause with regard to the dissimilarity measure it interferes with the calculation of the cluster centers. Cluster centers are normally represented by the cluster means in the K-means++ clustering algorithm. However, since not all values of the price elasticities are available we can not simply acquire the cluster mean. Therefore, the original K-means++ clustering algorithm does not suffice and the assignment of new cluster centers in the Restricted K-means++ clustering algorithm proposed in this paper should be conducted differently. The cluster center is found with regard to the following optimization criterion:

$$c_z = \operatorname*{argmin}_{y \in S_z} \frac{1}{|S_z|} \sum_{s \in S_z} D(s,y), \tag{4.10}$$

where $S_z$ is the set of stores assigned to the cluster with label $z$ and $D(s,y)$ represents the dissimilarity between store $s$ and $y$. This means an existing store serves as a cluster center. Most importantly note that K-means++ can yield empty clusters due to the assignment of cluster means. However, since our method does not simply compute the cluster means, but according to Equation 4.10 returns the store which on average has minimum distance to the other stores within a cluster, we prevent this from happening. The algorithm of the K-means++ clustering algorithm with a modification to the process of finding cluster centers can be found in Algorithm 1.

**Algorithm 1** K-means++

Initialize $k_g$ the number of clusters for PG $g$
Initialize $S$ the set of stores
Initialize $iterMax$ the number of iterations of K-means
Pick first initial center $c_1$ uniformly at random from $S$
**for** $l = 1$ **to** $k_g - 1$ **do**
    Pick new center $c_{l+1}$ from $S$ with probability $P(c_{l+1} = y) = \frac{\min_{c \in C_l} D(y,c)^2}{\sum_{s \in S} \min_{c \in C_l} D(s,c)^2}$
**end for**
**for** $n = 1$ **to** $iterMax$ **do**
    **for** $z = 1$ **to** $k_g$ **do**
        **if** $s \in S$ is closest to center $c_z$ **then**
            Assign $s$ to cluster with center $c_z$
        **end if**
        Set new cluster center $\underset{y \in S_z}{\operatorname{argmin}} \frac{1}{|S_z|} \sum_{s \in S_z} D(s, y)$
    **end for**
**end for**

*Retailer specifics*
To define the clusters retailers operate in compliance with domain-specific business rules. One of the these business rules that we should take into account during clustering is that stores originating from the same domicile should be assigned to the same cluster. This makes us propose the Restricted K-means++ clustering algorithm as found in Algorithm 2. To alter the original clustering algorithm for domain-specifics, the constraint should be handled with care during initialization or assignment [10]. Since we can easily depict manually which stores should be assigned to the same clusters according the aforementioned geographical business rule a simple adaption of the assignment of stores should suffice. Assume $V$ is the set of stores located in the same domicile. Inspired by Wagstaff et al. [40], assign store $v$, $\forall v \in V$, to the cluster with center $c_z$ according the following optimization criterion:

$$c_z = \underset{c \in C}{\operatorname{argmin}} \frac{1}{|V|} \sum_{v \in V} D(v, c). \tag{4.11}$$

Hence, Equation 4.11 represents the average distance of the stores located in the same domicile to the closest cluster center. Thus, all stores located in the same domicile are assigned to a cluster if these stores are closest on average to the center of that cluster. This means at first all stores are separately clustered and subsequently cluster centers are determined as proposed in Equation 4.10. Last, taking into account the business rule, stores are re-assigned according Equation 4.11.

*Robustness*

Additionally, we focus on building a robust system by conducting the Restricted K-means++ clustering $nStart$ times taking the clustering $r^*$ that minimizes the total within sum-of-squares (TWSS) of the clusters. The TWSS is a measure of internal cohesion and is the sum of distance functions for all clusters of each point in the cluster to the $k^{th}$ center. This way we avoid that little bit of randomness that slips into the system by picking a first cluster center uniformly at random during random seeding.

**Algorithm 2** Domain-specific Restricted K-means++

Initialize $k_g$ the number of clusters for PG $g$
Initialize $S$ the set of stores
Initialize $iterMax$ the number of iterations of K-means
Initialize $V$ the set of stores per domicile
Pick first initial center $c_1$ uniformly at random from $S$
**for** $l = 1$ **to** $k_g - 1$ **do**
    Pick new center $c_{l+1}$ from $S$ with probability $P(c_{l+1} = y) = \frac{\min_{c \in C_l} D(y,c)^2}{\sum_{s \in S} \min_{c \in C_l} D(s,c)^2}$
**end for**
**for** $n = 1$ **to** $iterMax$ **do**
    **for** $z = 2$ **to** $k_g$ **do**
        **if** $s \in S$ is closest to center $c_z$ **then**
            Assign $s$ to cluster with center $c_z$
        **end if**
        Set new cluster center $\underset{y \in S_z}{\operatorname{argmin}} \frac{1}{|S_z|} \sum_{s \in S_z} D(s, y)$
    **end for**
    **if** $V \subset S$ closest to center $c_z$ subject to $c_z = \underset{c \in C}{\operatorname{argmin}} \frac{1}{|V|} \sum_{v \in V} D(v, c)$ **then**
        Assign $v$, $\forall v \in V$ to cluster with center $c_z$
    **end if**
**end for**

Note, the Restricted K-means++ clustering algorithm is conducted for all product groups that are considered separately.

### 4.2.2 Cluster consensus

As mentioned before, the behavior among product groups differs and therefore we found different clusters of stores for each product group separately. However, in the end the retailer searches for one specific clustering of stores due to operational and optimization limitation. The information in the separate clusterings of product groups is captured in a single relation by means of the hard least squares Euclidean consensus algorithm [17] through the CLUE package in open source software package R.

Before the actual clustering consensus can be found the definition of dissimilarities between clusterings should be considered. The main issue to calculating dissimilarities between clusterings is that the cluster labels may be mixed up in different clusterings, whereas the underlying structure of the data does not change. Therefore we have to find a permutation of the cluster labels in such a way that agreement between clusters is maximized. In order to assess the problem of cluster labels permutation matrix $\Pi$ is used to replace $\lambda$ by $\lambda\Pi$. This relabeling procedure assures the cluster labels of $\lambda$ are reassigned without changing the underlying structure of the data. Since searching through all possible permutations is very time-consuming Dimitriadou et al. [8] defined the following Euclidean clustering dissimilarity between $\lambda$ and $\widetilde{\lambda}$:

$$d(\lambda, \widetilde{\lambda}) = \min_{\Pi} \|\lambda - \widetilde{\lambda}\Pi\|, \tag{4.12}$$

where $\|\cdot\|$ is the Frobenius norm. This distance measure is simply equivalent to maximizing the number of objects with the same class id in both clusterings. For proof see Hornik [17]. Now, Hornik proposed to find the optimal $\Pi$ using the Hungarian method [25] since we have to deal with the linear sum assignment problem (LSAP). Subsequently, all information in the elements of the cluster ensemble $\Lambda$ needs to be incorporated into the so-called cluster consensus. The cluster consensus is the single clustering of stores we are looking for. Consensus candidates $\lambda$ are compared to the cluster ensemble. The optimal cluster consensus $\lambda^*$ is defined as follows:

$$\lambda^* = \operatorname*{argmin}_{\lambda} \sum_{\lambda_g \in \Lambda} x_g d(\lambda, \lambda_g)^p, \tag{4.13}$$

which is equivalent to

$$\lambda^* = \operatorname*{argmin}_{\lambda} \sum_{\lambda_g \in \Lambda} x_g \min_{\Pi_g} \|\lambda - \lambda_g \Pi_g\|^p., \tag{4.14}$$

where $x_g$ is the PG-specific weight. Moreover, weights are based on the revenue of a specific product group:

$$x_g = \frac{\sum_{j \in P_g} \overline{\psi_j} d_j}{\sum_{h=1}^{G} \sum_{j \in P_h} \overline{\psi_j} d_j}. \tag{4.15}$$

Then, $x_g$ represents the total revenue of product group $g$ divided by the total revenue of all product groups such that $0 \leq x_g \leq 1$. If $p = 2$, we use least squares consensus clusterings. Now define $\widetilde{\lambda}$ as the weighted average of $\lambda_g \Pi_g$ and fix $\Pi_g$. The weighted average of $\lambda_g \Pi_g$ is equivalent to:

$$\widetilde{\lambda} = \frac{1}{\sum_{g=1}^{G} x_g} \sum_{g=1}^{G} x_g \lambda_g \Pi_g. \tag{4.16}$$

Then, Hornik. [17] proves the optimal $\lambda$ which is $\lambda^*$, is given by $\widetilde{\lambda}$. Moreover, since we have to find multiple optimal permutations for the separate partitions we no longer assess the LSAP but it becomes a multi-dimensional assignment problem (MAP) which is solved using the DWH (Dimitriadou, Weingessel and Hornik) algorithm from the CLUE package in R. This algorithm is an extension of the greedy algorithm described in Dimitriadou et al. [8]. Note that without going back to the information contained in the original data or the Restricted K-means++ clustering algorithm we obtain a final clustering.

### 4.2.3 Cluster validation

In general there are two different validity tests to determine the number of clusters that fits the underlying data the best: internal indices versus external indices [42]. Internal indices are used to measure the goodness of a clustering structure without respect to external information [37]. On the contrary, external indices are used when cluster labels are known a priori [9]. For our purpose we use internal indices as cluster labels are unknown beforehand. A lot of research is conducted for cluster validity from which we selected the most commonly applied elbow method [31] as a decision rule to determine $k_g$ the number of clusters for product group $g$[1]. The number of clusters used for the cluster consensus $k^*$ is found by using the weighted average of the number of clusters found for the separate product groups:

$$k^* = round(\sum_{g=1}^{G} x_g k_g), \tag{4.17}$$

where $round(\cdot)$ is the nearest integer function, $x_g$ is the weight for product group $g$ and $k_g$ the number of clusters indicated for product group $g$.

### 4.2.4 Cluster evaluation

The price policies applied in this paper embody a simplified version of reality where no price optimization method comes in play. The main focus is proving usefulness of our clustering framework by searching for the impact of the acquired clusters of stores and not so much to find optimal prices.

The main goal of the price setting procedure is to create price differences for the products between the clusters. The price setting process comprises two steps: label products as elastic, regular, or inelastic and subsequently change prices of elastic and inelastic labeled products. Therefore, we choose to label products based on predefined product-specific intervals. The average price elasticity $\mu_i$ and standard deviation $\sigma_i$ of product $i$ are found and used as input to construct the product-specific interval $(\mu_i \pm \sigma_i)$, see Figure 4.2.

[1]The silhouette index [35] and the Hartigan index [15] are used for confirmation. In case both the silhouette index and Hartigan index do not agree on the number of clusters $k_g$ found by the elbow method for product group $g$, the number of clusters is reconsidered.

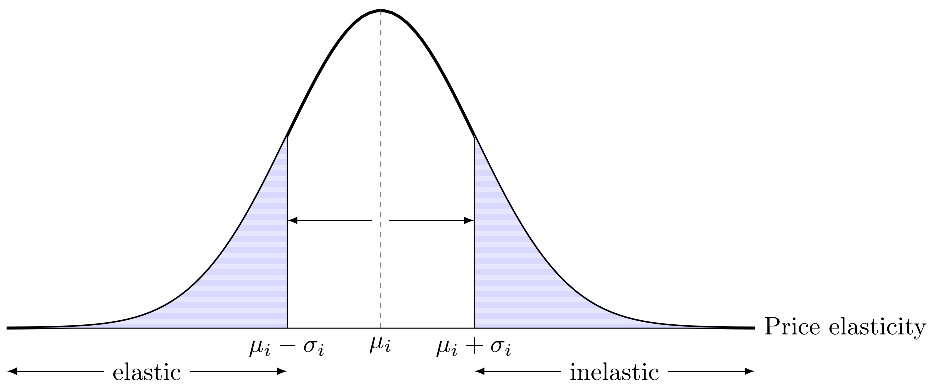


**Figure 4.2:** The illustration of the product-specific interval.

Next, labels are assigned to each product within a specific cluster. The average price elasticity $\mu_{i,c}$ of product $i$ in cluster $c$ is found and compared to the product-specific interval. In case $\mu_{i,c}$ exceeds the lower bound of the interval and $\mu_{i,c}$ is smaller than -1 the product is assigned the label 'elastic'. The cluster-specific average price elasticity of a specific product should exceed -1, otherwise a product is in principle not elastic at all. Whenever $\mu_{i,c}$ exceeds the upper bound of the interval and $\mu_{i,c} \in (-1, 0)$ the product is assigned the label 'inelastic'. In all other cases the product is assumed 'regular'. The labeling procedure is summarized in Algorithm 3.

**Algorithm 3** Labeling procedure

Initialize $P_g$ the set of products in a PG
Initialize $C$ the set of cluster centers
**for** $i \in P_g$ **do**
  Find average $\mu_i$ and st.dev. $\sigma_i$ of the price elasticity of product $i$
  **for** $c \in C$ **do**
    Find average $\mu_{i,c}$ of the price elasticity of product $i$
    cluster with cluster center $c$
    **if** $\mu_{i,c} < \mu_i - \sigma_i$ **and** $\mu_{i,c} < -1$ **then**
      Label product $i$ in cluster with cluster center $c$ as 'elastic'
    **else if** $\mu_{i,c} > \mu_i + \sigma_i$ **and** $-1 < \mu_{i,c} < 0$ **then**
      Label product $i$ in cluster with cluster center $c$ as 'inelastic'
    **else**
      Label product $i$ in cluster with cluster center $c$ as 'regular'
    **end if**
  **end for**
**end for**

The following step is to set prices for a new price policy based on the labeling procedure. Current prices are used as a starting point. Obviously, the prices of inelastic labeled products are raised and prices of elastic labeled products are cut. Since we have the data of Jumbo at our disposal we have to take into account their business rules. Therefore, prices should not exceed the lowest prices of the competition within the same domicile. This leads to a price increase in inelastic products up to the prices of the competition with a maximum of 5%. The price decrease of elastic products is limited to a maximum of 5%. Both the percentage changes are set according the policy of the retailer. Last, price elasticities are used to find the projected sales after the price setting procedure is conducted. Keep in mind that we force price differences between clusters in order to be able to prove the value to the retailer to set different prices in different clusters of stores.

To conduct a robustness analysis for the clustering we use the Jaccard (dis)similarity [20]. The Jaccard similarity is developed to find the similarity between sets of categorical variables, such as the similarity between the cluster consensus and the separate clusterings of each product group. This measure is specifically chosen since the distances between the cluster labels which represent cluster membership are irrelevant. In Equation 4.19 the formula of the Jaccard similarity can be seen, where $\lambda_g(S)$ and $\lambda^*(S)$ are the clustering for product group $g$ and the cluster consensus representing cluster membership:

$$J(\lambda_g(S), \lambda^*(S)) = \frac{|\lambda_g(S) \cap \lambda^*(S)|}{|\lambda_g(S) \cup \lambda^*(S)|} \tag{4.18}$$

where $\lambda_g(S) \cap \lambda^*(S)$ = the intersection of sets $\lambda_g(S)$ and $\lambda^*(S)$ and $\lambda_g(S) \cup \lambda^*(S)$ = the union of sets $\lambda_g(S)$ and $\lambda^*(S)$. The opposite of the Jaccard simmilarity is obviously the Jaccard dissimilarity which is as follows:

$$J(\lambda_g(S), \lambda^*(S)) = 1 - \frac{|\lambda_g(S) \cap \lambda^*(S)|}{|\lambda_g(S) \cup \lambda^*(S)|}. \tag{4.19}$$

Below we give an example of the (dis)similarity calculation between the cluster membership for a specific product group $\lambda_g(S)$ and the cluster consensus $\lambda^*(S)$. Imagine hypothetically we have 8 different stores assigned to two different clusters.

$$\begin{aligned} \lambda_g(S) &= \begin{bmatrix} 2 & 1 & 1 & 2 & 1 & 2 & 2 & 2 \end{bmatrix}^T \\ \lambda^*(S) &= \begin{bmatrix} 1 & 1 & 1 & 2 & 2 & 2 & 2 & 1 \end{bmatrix}^T \end{aligned} \tag{4.20}$$

It can be easily seen that the Jaccard similarity equals 5/8 and the Jaccard dissimilarity equals 3/8. This means that for the evaluated product group 5 out of the 8 stores end up in the same cluster as in the cluster consensus.

## 4.3 | Regression analysis

Next, we acquire the drivers of price elasticity by conducting regression models for each of the separate clusters. This way we can prove the differences in consumer behavior among the clusters and gain insight in the underlying factors that affect consumer behavior such the retailer can make better decisions. In order to discover the drivers of differences in consumer behavior among clusters panel data models [28] are conducted for the clusters, inspired by Hoch et al. [16]. Therefore socio-demographic as well as socio-economic variables and competitional characteristics among others are used [26]. In Subsection 4.3.1 the theoretical background on regression models is elaborated. Subsection 4.3.2 explains the variables that are made available and discusses the selection procedure of these variables for the regression models.

### 4.3.1 Regression model

We consider a unique regression model for each of the four different clusters. The regression model serves descriptive purposes [18] for the price elasticity of stores per cluster. So, in each of these models the dependent variable is the price elasticity per store. For each regression model the price elasticities per store of the stores that are assigned to the considered cluster are used. Since time contributes to the diversity of the data we consider panel data regression models for each of the formed clusters. We use time-variant as well as time-invariant characteristics. The step wise decision process to acquire the final regression model leads us to the considered random effects models. The following model postulates the relationship between price elasticities per store per cluster, $e_{s,t}$, and covariates, $x_{s,t}$, as follows:

$$e_{s,t} = \delta + x'_{s,t}\gamma + u_{s,t}, \tag{4.21}$$

with $u_{s,t} = (\delta_s - \delta) + \eta_{s,t}$, $\delta_s \sim i.i.d.(\delta, \sigma^2_\delta)$, and $\eta_{s,t} \sim i.i.d.(0, \sigma^2_\eta)$. The price elasticities per store per cluster are found by aggregating the sales of all products part of all six considered product groups in a specific store. Note, the price elasticities considered in the regression models are no longer per product per store since we are especially interested in the drivers that separate the stores and not so much in product specifics. The feasible generalized least squares estimation (FGLS) is used to estimate the parameters in the random effects model.

### 4.3.2 Variable selection

The extensive background information on store-specifics and trading area comprises socio-economic variables, socio-demographics and competitive characteristics among others. An overview of all the variables used during the regression can be found in Table A.2 in the Appendix. We conduct a step wise regression with forward selection. To counterbalance the problem of multiple comparisons we need to assure the family-wise error rate (FWER) is controlled at the significance level ($FWER \leq \alpha$). The controlling procedure we use to encounter this problem is the classical Bonferroni correction [6]:

$$p_i \leq \frac{\alpha}{m} \tag{4.22}$$

where $p_i$ is the newly found significance level for hypothesis $i$, $\alpha$ is the desired significance level, and $m$ is the number of hypotheses tested. In general, the significance level $\alpha$ is set to 5%.

With regard to the time-dependent variables we have to deal with aggregation issues. Since the data is made available at store-item level per week we need to aggregate over the products sold in a store to acquire store level data per week. Moreover, for the prices of substitutes and the prices of complements the average price during a week is used. For the promotional variables the percentage of products on promotion is composed.

# 5 | Evaluation

In this chapter the results of the proposed methods are discussed. First, the cluster definition is elaborated in Section 5.1. Next, in Section 5.2 the results of the sensitivity analysis for different price policies are considered. In Section 5.3 we discuss the results of the new price policy. Finally, in Section 5.4 we take a closer look at the results from the regression analysis.

## 5.1 | Cluster definition

This section provides the cluster definition. We show the cluster validation of the clustering algorithm and the interpretation of each of the found clusters in Subsection 5.1.1. Next, we evaluate the robustness of the clustering method using the Jaccard dissimilarity in Subsection 5.1.2.

### 5.1.1 Cluster interpretation

The number of clusters indicated by the elbow method for each of the separate product groups is presented in the elbow plot in Figure 5.1 by means of the TWSS[1]. Results show to remain four clusters for each of the product groups. Therefore, we set the number of clusters for the cluster consensus to four as well.

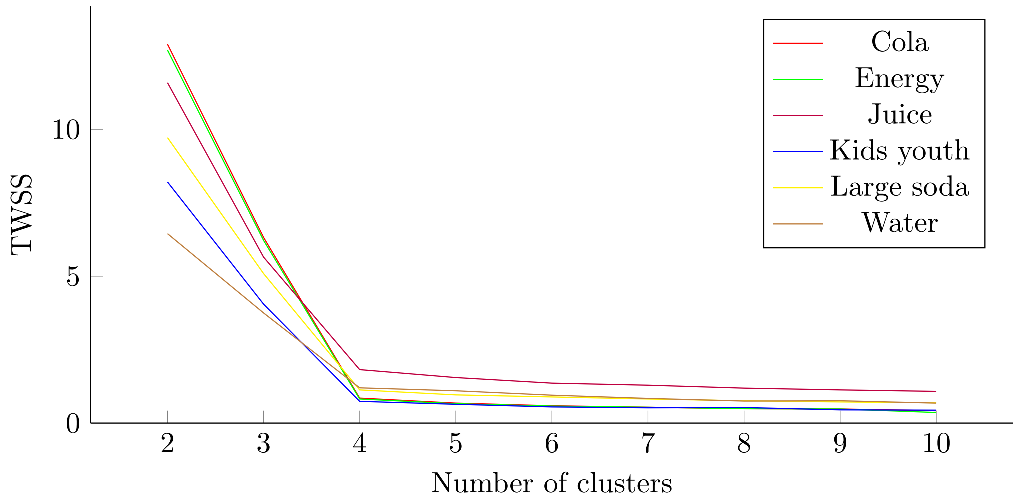


**Figure 5.1:** The figure presents the number of clusters per product group indicated by the elbow method by means of the total within sum of squares (TWSS).

The first cluster contains 102 stores represented by an environment with a relatively high percentage of consumers part of the low social class, see Table A.3 in the Appendix for

[1]The silhouette and Hartigan index confirm the number of clusters indicated by the elbow method.

more information on social classes. Moreover, the percentages of youth and non-western immigrants are large. The second cluster consists of 105 stores mostly situated in the South of the country. As well as for the fourth cluster, which consists of 41 stores, the percentage of consumers part of the high social class and families with kids are relatively high. These clusters differ in particular in the share of youth which is highest in cluster 4. Last, the remaining 57 stores belong to cluster 3 and are mostly situated in the upper North of the country in the rural areas which is reflected in the small share of immigrants and the fact that supermarkets are relatively far apart. As well as for cluster 1 the percentage of consumers part of the low social class is relatively high. An overview of the cluster means of some of the characteristics of the different clusters can be found in Table 5.1.

**Table 5.1:** The table presents the cluster means of some of the characteristics of the different clusters (in percentages).

| Characteristic | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| Low social class | 54 | 46 | 60 | 45 |
| High social class | 46 | 54 | 40 | 55 |
| Non-western immigrants | 8 | 5 | 4 | 5 |
| Western immigrants | 8 | 10 | 5 | 9 |
| Youth | 9 | 7 | 7 | 11 |
| Families with kids | 17 | 20 | 15 | 20 |
| Average distance to supermarket (in m) | 500 | 590 | 685 | 530 |

### 5.1.2 Cluster evaluation: Jaccard

Throughout Subsection 4.2.4 we have discussed the Jaccard dissimilarity. This measure is used as an indication of the robustness of the clustering method. Note, the results provided in Table 5.2 are measured with regard to the cluster consensus. This way each of the dissimilarities found in the table represents a comparison between the cluster membership of the listed product group and the cluster consensus.

**Table 5.2:** The Jaccard dissimilarities of the cluster memberships of the product groups versus the cluster consensus.

| Clusters | Cola | Energy | Juices | Kids youth | Large soda | Water |
|---|---|---|---|---|---|---|
| All stores | 0.039 | 0.026 | 0.026 | 0.020 | 0.007 | 0.026 |
| Cluster 1 | 0.056 | 0.064 | 0.019 | 0.057 | 0.019 | 0.073 |
| Cluster 2 | 0.076 | 0.028 | 0.057 | 0 | 0 | 0.019 |
| Cluster 3 | 0.070 | 0 | 0.035 | 0.068 | 0 | 0.070 |
| Cluster 4 | 0.128 | 0.146 | 0.128 | 0.049 | 0.049 | 0.049 |

In general, the Jaccard dissimilarity of the product groups with the cluster consensus seems strikingly low with 0.039 as a high for Cola. This indicates an overall difference of 12 out of 305 stores appointed to different clusters. This means the deviation of each of the clusterings of the product groups from the cluster consensus is small and the stores for the most part end up in the same cluster. Noteworthy is for example the fact that for both Kids youth and Large soda the stores in the second cluster do not differ from the cluster consensus. This could be due to the large number of products together with the relatively small spread in price elasticity in these product groups, see Table 3.1. This makes clustering relatively easy compared to the other product groups. Considering the results of the Jaccard dissimilarity the proposed Restricted K-means++ clustering method is very robust and only indicates minor differences between the clusterings of the product groups versus the cluster consensus. The consumer behavior seems to be split roughly the same for each of the product groups since we are using the most specific level of price elasticities.

## 5.2 | Sensitivity analysis

In Figure 5.2 we provide the initial average revenue and profit per product of the evaluated product groups as a reference point for the upcoming results. The average revenue and profit are based on the current out-of-the-door prices, also called regular prices. The average revenue as well as the average profit are at a maximum for the product group Cola while these numbers are at a minimum for the product group Kids youth. Cola can be seen as one of the most popular product groups since the sales of the individual products are relatively high. From Figure 5.2 it follows the average profit share per product is 27%. That is why the average profit share per product of 32% for Energy stands out since the margin per product is relatively high.

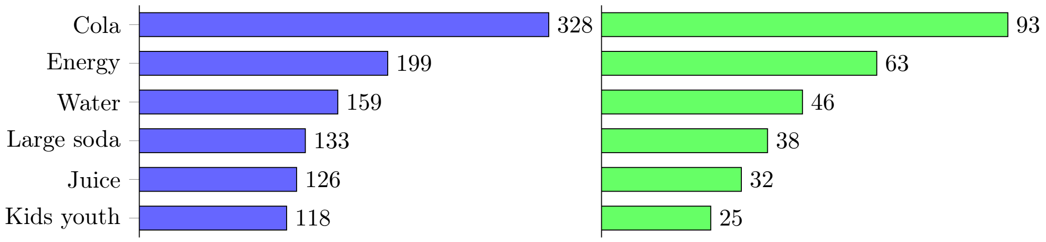


**Figure 5.2:** The initial situation of the average revenue (l) and average profit (r) per product of the products within a product group (in thousands of euros).

We work step by step towards a new price policy as discussed in Subsection 4.2.4. After each product is indicated as being ‘elastic’, ‘inelastic’ or ‘regular’, new prices are set for the different clusters. To recap, the prices of inelastic products are increased up to a maximum of 5% and the price cut of elastic products is also limited to 5%. Moreover, price elasticities are used to find the projected sales after new prices are set. If we change prices we have to reckon with competition since prices are not supposed to exceed the prices of the competition and Jumbo has the ‘lowest price guarantee’, which basically means that prices of a specific product for all stores in a cluster are set as low as necessary by the supermarket in a way that there is still sufficient profit left.

First, we take a closer look at the effect of price changes of elastic products on the revenue and profit. Obviously, by increasing the percentage change of prices of both the elastic and inelastic products the revenue increases. On the other hand this is not by definition the case considering the profit. Results show that cutting prices of elastic products by a predefined percentage in general is not profitable in the end. The reason is that prices are bounded to the business rule of the lowest price guarantee. There is no longer space for price cuts without harming the profit due to the negligible difference between the cost price and out-of-the-door price in general. If we cut prices it does not lead to sufficient increases in the sales such the profit grows.

Now, we take a closer look at the the effect of price changes of inelastic products on the revenue and profit. From now on, we consider a maximum price increase in inelastic products, indicated by ‘max+5%’, as the price increase which is bounded by the price of that product charged by the competition within the same domicile. For example, product $i$ and $j$ are priced €1 in our store. The price of the competition for product $i$ is €1,05 and the price of the competition for product $j$ is €1,10. The price of product $i$ can be raised up to €1,04, since the new price can not exceed the price of the competition. This results in a price increase of 4%. For product $j$ we can raise the price up to €1,09 which represents a price increase of 9%. However we are bounded to a maximum increase of 5% of the current price resulting in a new price of €1,05 for product $j$. Results show that changing prices of inelastic products results in more revenue and more profit in the end.

In conclusion, a price change of elastic products in general does not lead to an increase in profit, but price changes of elastic products prove to be valuable in case of the revenue. On the other hand, the price increase in inelastic products is valuable to the revenue and is profitable. Thus, the optimal profit can be found by not changing prices of elastic products and changing prices of inelastic products up to the maximum of +5%. In the end we are dealing with a trade-off between revenue and profit. The increase in revenue can be related to more customers and therefore more relevance of the retailer in the eyes of the customer. On the other hand profit is important for the retailer itself. The impact on the revenue and profit by means of the step wise percentage change in price of inelastic and elastic products can be found in Table B.1 and B.2 in the Appendix, respectively.

**Table 5.3:** The projected increase in revenue and profit for different price policies (in percentages).

| Elastic | Inelastic | Revenue | Profit |
|---|---|---|---|
| -0% | max+5% | 0.19 | 1.1 |
| -1% | max+5% | 0.28 | 0.93 |
| **-2%** | **max+5%** | 0.36 | 0.76 |
| -3% | max+5% | 0.45 | 0.57 |
| -4% | max+5% | 0.52 | 0.37 |
| -5% | max+5% | 0.6 | 0.15 |

Table 5.3 presents the sensitivity analysis for the projected increase in revenue and profit for different price policies in percentages. As mentioned before, we noticed that changing prices of elastic products results in an increase in revenue and in a decrease in profit generally. However, looking at price changes of inelastic products both revenue and profit result in an increase. We impose a trade-off between revenue and profit. To find middle ground, from this moment onward we focus on the price policy where prices of elastic products are cut by 2% and prices of inelastic products are raised by max+5%. In this case, the projected increase in revenue and profit of the considered product groups are 0.36% and 0.76%, respectively.

## 5.3 | New price policy

Now, we are going to highlight some of the results of the new price policy discussed in Section 5.2. Tables 5.4 and 5.5 present the projected increase in revenue and profit per product group by cluster, respectively. Note, we are taking a closer look at the projected *increase* in revenue and profit and not at the total effective absolute values of revenue and profit. Thereby it can happen that the increase in revenue exceeds the increase in profit which could seem counter intuitive. This has everything to do with the co-existence of the change in sales and change in total cost both found using the price elasticity. The moment we set new prices according the price elasticity, the sales change and therefore the total cost of these products change.

**Table 5.4:** The projected increase in revenue per product group by cluster with competition taken into account and 2% price change of elastic products.

| Product group | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Total |
|---|---|---|---|---|---|
| Cola | 18,528 (0.15) | 76,274 (0.63) | 23,222 (0.39) | 10,505 (0.20) | 128,529 (0.36) |
| Energy | 531 (0.01) | 3,605 (0.08) | 13,626 (0.60) | 6,506 (0.49) | 24,268 (0.18) |
| Juice | 7,960 (0.08) | 47,689 (0.52) | 38,157 (0.62) | 18,141 (0.43) | 111,947 (0.38) |
| Kids youth | 32,077 (0.34) | 36,784 (0.39) | 40,109 (0.63) | 9,693 (0.32) | 118,663 (0.42) |
| Large soda | 9,953 (0.12) | 54,504 (0.53) | 34,461 (0.61) | 8,973 (0.32) | 107,891 (0.40) |
| Water | 6,692 (0.14) | 17,170 (0.29) | 16,120 (0.65) | 8,312 (0.44) | 48,294 (0.32) |
| Total | 75,741 (0.15) | 236,026 (0.46) | 165,695 (0.58) | 62,130 (0.34) | 539,772 (0.36) |

Table 5.4 gives the projected increase in revenue per product group by cluster. In line with our expectations the relative increase in revenue of cluster 3 is highest compared to the other clusters by reason of the substantial share of products indicated as elastic or inelastic, namely 19.2%. However, in absolute terms cluster 2 outperforms cluster 3 since it consists of almost double the number of stores. The average number of products indicated as inelastic and elastic per cluster is 17.5%. For more information on the number of products indicated as elastic or inelastic see Table B.4 in the Appendix. In addition, outstanding is the minor increase in revenue for cluster 1 despite the large amount of stores in cluster 1. The number of products indicated as inelastic (6%) is strikingly low. Further notice that Kids youth experiences the most growth in relative terms and in particular cluster 1 stands out with regard to the other product groups.

**Table 5.5:** The projected increase in profit per product group by cluster with competition taken into account and 2% price change of elastic products.

| Product group | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Total |
|---|---|---|---|---|---|
| Cola | 27,256 (0.79) | 9,807 (0.28) | 23,112 (1.34) | 2,512 (0.18) | 62,687 (0.63) |
| Energy | -71 (-0.00) | -714 (-0.01) | 16,238 (2.13) | -2,319 (-0.56) | 13,134 (0.31) |
| Juice | 4,869 (0.19) | 21,703 (0.94) | 29,692 (1.92) | 69 (0.00) | 56,333 (0.76) |
| Kids youth | 18,063 (0.92) | 41,332 (2.01) | 2,176 (0.16) | 4,320 (0.67) | 65,890 (1.10) |
| Large soda | 11,878 (0.50) | 28,838 (1.01) | 35,322 (2.15) | 428 (0.01) | 76,465 (0.99) |
| Water | -1,022 (-0.07) | 12,973 (0.78) | 10,984 (1.49) | 2,747 (0.51) | 25,683 (0.59) |
| Total | 60,973 (0.46) | 113,939 (0.83) | 117,524 (1.52) | 7,757 (0.16) | 300,193 (0.76) |

Then, the projected increase in profit per product group by cluster is depicted in Table 5.5. In general the different clusters of each of the product groups are more profitable than the initial situation. Nonetheless, we can find some clusters, such as cluster 2 of the product group Energy and cluster 1 of the product group Water, which result in a decrease in profit. Especially for cluster 2 of the product group Energy this is not surprising since exclusively some products are indicated as elastic and as we have seen before price changes of elastic products usually result in a decrease in profit. Cluster 3 outperforms cluster 2 although the projected increase in revenue for cluster 2 is higher. This indicates the cost of goods sold changes in favor of cluster 3. In absolute terms as well as in relative terms cluster 4 is least profitable among other things because of the small number of stores that end up in cluster 4. Then, the projected increase in profit in relative terms is highest for the product group Kids youth. Likewise, the projected increase in profit of the product group Kids youth is highest.

Finally, results show a projected increase in revenue of 0.36% and a projected increase in profit of 0.76% in total for the considered product groups. Cluster 3 is most profitable as a result of the large amount of inelastic indicated products (11%) despite the fact that cluster 2 generates a higher projected increase in revenue. With regard to the product groups, especially the profit increase in Kids youth attracts attention endorsed by the projected increase in revenue.

## 5.4 | Regression analysis

At last we are interested in the drivers of consumer behavior and thereto this section discusses the results of the panel data regressions[2]. To recall regression analysis is conducted such we can prove the differences in consumer behavior among the clusters and gain insight in the underlying factors that affect consumer behavior such the retailer can make better decisions. Note, when interpreting the signs of the coefficients of the independent variables it is important that the dependent variable is a negative number. If the independent variable grows, a positive sign results in a shrinking effect on the price elasticity, whereas a negative sign indicates an enlarging effect.

Table 5.6 presents a few of the covariates of the price elasticities per store per cluster. First, the holidays Father's and Mother's day are considered which show various effects on the price elasticity. Father's day has an enlarging effect on the price elasticity, whereas Mother's day has a shrinking effect. The considered advertisement channels in this section comprise instore flyers and national folders since they prove the difference in behavior among clusters. Generally, these advertisement channels increase price sensitivity, however in particular cluster 3 behaves differently. The price of substitutes seems to have an enlarging effect in general on the price elasticity, however for cluster 2 this is not the case. Interesting is the fact that time-independent variables usually have significant effect on the price elasticities in exclusively one of the clusters. Outstanding is the difference in effect of unemployment on clusters 1 and 2. The unemployment rate results in an increasing effect of price sensitivity for cluster 1 and a decreasing effect for cluster 2. Summarizing, these covariates provide sufficient evidence to assume that consumer behavior among clusters differs and therefore pricing decisions should be adjusted accordingly. The full

[2]Linear time-independent regressions were also attempted but yielded no significant results.

results of the panel data regression models with regard to the clusters can be found in Table C.8 in the Appendix. Additionally, the panel data regression models per product group per cluster are also elaborated in Appendix C to show differences within each product group between the four found clusters. Furthermore, this Appendix contains the results of regression models for the different product groups to show differences in consumer behavior among product groups.

**Table 5.6:** The table presents some of the coefficients (and standard errors) of the variables of the panel data regression models on store-item level for all clusters to declare the dependent variable of price elasticities. Note, that the significance level is set to 5% and only variables indicated by '-' are not significant.

| **Panel data regression model**<br>**Independent variables** | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| Father's day | -0.025 (0.004) | -0.033 (0.005) | -0.027 (0.005) | -0.027 (0.008) |
| Mother's day | 0.050 (0.006) | 0.054 (0.007) | 0.029 (0.007) | 0.069 (0.012) |
| Price ratio substitute 1 | -1.083 (0.039) | 0.316 (0.033) | -1.183 (0.049) | -0.387 (0.064) |
| Instore flyer | -0.446 (0.12) | -0.469 (0.139) | 0.000 (0.000) | -0.786 (0.247) |
| National folder | -1.203 (0.140) | -0.640 (0.120) | 0.000 (0.000) | -0.551 (0.216) |
| Percentage high education | -0.105 (0.035) | - | - | - |
| Percentage unemployed | -0.410 (0.202) | 0.789 (0.202) | - | - |
| Percentage regular visitor competitor | -0.050 (0.023) | - | - | - |
| $R^2$ | 0.622 | 0.649 | 0.651 | 0.802 |

# 6 | Conclusion

In this paper we proposed a two-stage clustering framework using differences in consumer behavior to define new clusters of stores which are not solely based on local competition. This way we can boost the revenue and profit of a retailer. The context used is one of the major supermarket chains in the Dutch market. We defined clusters of stores within the already existing clusters of stores as there appears to be a difference in consumer behavior among stores within product groups reflected in the variance in price elasticities. First, we proposed the Restricted K-means++ clustering algorithm to find the clusterings of stores for each product group separately. Retailer specifics are implemented under the guise of a geographical restriction such that each store situated in the same domicile is appointed to the same cluster. Second, we combined the information comprised in these separate clusterings into a general clustering called the cluster consensus. That way, all information on the price elasticity at store-item level is used and no information iss lost in the process due to aggregation issues. In the following, we highlight our contributions and propose suggestions for future work.

First, the clustering in general is very accurate and robust since the Jaccard dissimilarity indicates small differences between the clusterings of the product groups and the cluster consensus. Stores are by and large appointed to the same cluster within each of the product groups which indicates using price elasticities on store-item level seems an appropriate choice. Then, most retailers set different prices in clusters of stores based on the competition in proximity of the considered stores. Our research provides an easily implementable and fully automated clustering method to find groups of 'similar' stores which are not exclusively driven by local competition. The results show the influence of competition, store-specifics and trading area data on the price sensitivity, proving consumer behavior is driven by these factors on a day-to-day basis. Thus, depending on the extent of the influence of the aforementioned characteristics the pricing decision should be informed accordingly. Furthermore, the algorithm is not order-sensitive by reason of the random seeding technique. Moreover, we know that K-means can yield empty clusters however by constructing cluster centers as we propose in this research this can not happen. The algorithm is not limited to the singular data set used in this research. Furthermore, it is not even limited to the context of supermarkets. The only information really needed to apply the algorithm are the price elasticities on store-item level and the revenue per product per store which makes the algorithm transparent. The revenue per product per store is needed to find the product importance and NACT proposed in this paper. The NACT provides a convenient way to cope with not available values. Altogether, the sensitivity analysis for the considered product groups suggests potential of defining new clusters of stores based on consumer behavior enhanced by the projected increase in revenue of 0.36% and the projected increase in profit of 0.76%.

The follow-up step is to consider if and how the analysis could be extended to the complete store, which means taking all product groups present in the store in consideration. Moreover, in this research we focus exclusively on the first cluster and are limited to the results that belong to this cluster. Therefore, an opinion on operational feasibility should be formed concerning the extension to complete stores and all existing pricelines. In general, it seems obvious to reconsider the time span of pricing optimization as price sensitivity changes over time. For example, price elasticity differs from 'regular' weeks around the holidays proven by the panel data analysis. These days prices are optimized mostly 2-3 times per year whereas it could be interesting and profitable to update prices more regularly. Although outside the scope of this research, one could think about applying the method by Brombin et al. [7] to search for the covariates of price elasticity if interested in the drivers of the differences between clusters. This simulation method provides an elegant way to find the independent variables for models of price elasticity. This way the overall process of regression could be fully automated like the clustering algorithm without interference of the end-user selecting the variables by hand.

# Bibliography


[1] Allenby, G.M., Rossi, P.E.: A marginal-predictive approach to identifying household parameters. Marketing Letters 4(3), 227–239 (1993)

[2] Arthur, D., Vassilvitskii, S.: k-means++: The advantages of careful seeding. In: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. pp. 1027–1035. Society for Industrial and Applied Mathematics (2007)

[3] Basu, S., Banerjee, A., Mooney, R.: Semi-supervised clustering by seeding. In: In Proceedings of 19th International Conference on Machine Learning (ICML-2002. Citeseer (2002)

[4] Basu, S., Banerjee, A., Mooney, R.J.: Active semi-supervision for pairwise constrained clustering. In: SDM. vol. 4, pp. 333–344. SIAM (2004)

[5] Basu, S., Bilenko, M., Mooney, R.J.: A probabilistic framework for semi-supervised clustering. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 59–68. ACM (2004)

[6] Bonferroni, C.E.: Teoria statistica delle classi e calcolo delle probabilita. Libreria internazionale Seeber (1936)

[7] Brombin, C., Finos, L., Salmaso, L.: Adjusting stepwise p-values in generalized linear models. In: International Conference on Multiple Comparison Procedures.–see step. adj () in the R someMTP package (2007)

[8] Dimitriadou, E., Weingessel, A., Hornik, K.: A combination scheme for fuzzy clustering. International Journal of Pattern Recognition and Artificial Intelligence 16(07), 901–912 (2002)

[9] Dudoit, S., Fridlyand, J.: A prediction-based resampling method for estimating the number of clusters in a dataset. Genome biology 3(7), 1–21 (2002)

[10] Eduardo, M., Brea, A., et al.: Constrained clustering algorithms: Practical issues and applications (2013)

[11] Elrod, T., Winer, R.S.: An empirical evaluation of aggregation approaches for developing market segments. The Journal of Marketing pp. 65–74 (1982)

[12] Forgy, E.W.: Cluster analysis of multivariate data: efficiency versus interpretability of classifications. Biometrics 21, 768–769 (1965)

[13] Ghaemi, R., Sulaiman, M.N., Ibrahim, H., Mustapha, N., et al.: A survey: clustering ensembles techniques. World Academy of Science, Engineering and Technology 50, 636–645 (2009)

[14] Gordon, A.D.: Classification, (chapman & hall/crc monographs on statistics & applied probability) (1999)

[15] Hartigan, J.A.: Clustering algorithms (1975)

[16] Hoch, S.J., Kim, B.D., Montgomery, A.L., Rossi, P.E.: Determinants of store-level price elasticity. Journal of marketing Research pp. 17–29 (1995)

[17] Hornik, K.: A clue for cluster ensembles. Journal of Statistical Software 14(11) (2005)

[18] Hsiao, C.: Analysis of panel data. No. 54, Cambridge university press (2014)

[19] Huang, M.H., Hahn, D.E., Jones, E., et al.: Determinants of price elasticities for store brands and national brands of cheese. In: American Agricultural Economics Association Annual Meeting, Denver, Colorado. August. pp. 1–4 (2004)

[20] Jaccard, P.: The distribution of the flora in the alpine zone. New phytologist 11(2), 37–50 (1912)

[21] Karypis, G., Kumar, V.: Multilevelk-way partitioning scheme for irregular graphs. Journal of Parallel and Distributed computing 48(1), 96–129 (1998)

[22] Karypis, G., Kumar, V.: Parallel multilevel series k-way partitioning scheme for irregular graphs. Siam Review 41(2), 278–300 (1999)

[23] Kaufman, L., Rousseeuw, P.: Clustering by means of medoids. North-Holland (1987)

[24] Klein, D., Kamvar, S.D., Manning, C.D.: From instance-level constraints to space-level constraints: Making the most of prior knowledge in data clustering (2002)

[25] Kuhn, H.W.: The hungarian method for the assignment problem. Naval research logistics quarterly 2(1-2), 83–97 (1955)

[26] Kumar, V., Karande, K.: The effect of retail store environment on retailer performance. Journal of Business Research 49(2), 167–181 (2000)

[27] Lagin, M., Gebert-Persson, S.: Defining the links between retail price strategies and price tactics (2015)

[28] Maddala, G.S., Lahiri, K.: Introduction to econometrics, vol. 2. Macmillan New York (1992)

[29] Marn, M.V., Roegner, E.V., Zawada, C.C.: Pricing new products. McKinsey Quarterly (3), 40–49 (2003)

[30] Monroe, K.B.: Measuring price thresholds by psychophysics and latitudes of acceptance. Journal of Marketing Research pp. 460–464 (1971)

[31] Ng, A.: Clustering with the k-means algorithm. Machine Learning (2012)

[32] Petrick, J.F.: Segmenting cruise passengers with price sensitivity. Tourism Management 26(5), 753–762 (2005)

[33] Porro-Muñoz, D., Duin, R.P., Talavera, I.: Missing values in dissimilarity-based classification of multi-way data. In: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, pp. 214–221. Springer (2013)

[34] Rossi, P.E., Allenby, G.M.: A bayesian approach to estimating household parameters. Journal of Marketing Research pp. 171–182 (1993)

[35] Rousseeuw, P.J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics 20, 53–65 (1987)

[36] Strehl, A., Ghosh, J.: Cluster ensembles—a knowledge reuse framework for combining multiple partitions. Journal of machine learning research 3(Dec), 583–617 (2002)

[37] Thalamuthu, A., Mukhopadhyay, I., Zheng, X., Tseng, G.C.: Evaluation and comparison of gene clustering methods in microarray analysis. Bioinformatics 22(19), 2405–2412 (2006)

[38] Tumer, K., Agogino, A.K.: Ensemble clustering with voting active clusters. Pattern Recognition Letters 29(14), 1947–1953 (2008)

[39] Vega-Pons, S., Ruiz-Shulcloper, J.: A survey of clustering ensemble algorithms. International Journal of Pattern Recognition and Artificial Intelligence 25(03), 337–372 (2011)

[40] Wagstaff, K., Cardie, C., Rogers, S., Schrödl, S., et al.: Constrained k-means clustering with background knowledge. In: ICML. vol. 1, pp. 577–584 (2001)

[41] Wakefield, K.L., Inman, J.J.: Situational price sensitivity: the role of consumption occasion, social context and income. Journal of Retailing 79(4), 199–212 (2003)

[42] Wang, K., Wang, B., Peng, L.: Cvap: validation for cluster analyses. Data Science Journal 8, 88–93 (2009)

# Appendices

# A | Variables

**Table A.1:** The symbols used in the clustering method.

| Symbol | Description | Equation |
|---|---|---|
| $z$ | Cluster indicator | - |
| $n$ | Iteration indicator | - |
| $i, j$ | Product indicator | - |
| $r$ | Results indicator | - |
| $s, v, y$ | Store indicator | - |
| $t$ | Time indicator | - |
| $k$ | Number of clusters | - |
| $k_g$ | Number of clusters in PG $g$ | - |
| $k^*$ | Number of clusters in cluster consensus | - |
| $g$ | Product group indicator | - |
| $iterMax$ | Number of iterations K-means | - |
| $nStart$ | Number of iterations Restricted K-means++ | - |
| $G$ | Number of product groups | - |
| $I$ | Set of all products | 4.1 |
| $P_g$ | Set of all products sold in PG $g$ | 4.1 & 4.6 & 4.15 |
| $\Lambda$ | Set of final clustering results for each PG | 4.3 & 4.13 |
| $N(s, y)$ | Set of all products in a PG sold in store $s$ and $y$ | 4.4 & 4.5 & 4.7 & 4.8 |
| $S$ | Set of all stores | 4.9 |
| $C_l$ | Set of current cluster centers | 4.9 |
| $S_z$ | Set of stores assigned to the cluster with label $z$ | 4.10 |
| $V$ | Set of all stores located in the same domicile | 4.11 |
| $C$ | Set of cluster centers | - |
| $\Gamma$ | Set of input for the clustering of price elasticities on store-item level | - |
| $w_i$ | Weight of product $i$ | 4.4 & 4.5 & 4.8 |
| $\overline{\psi_i}$ | Average revenue per store for product $i$ | 4.4 & 4.6 & 4.7 & 4.15 |
| $d_i$ | Number of products in which product $i$ is sold | 4.4 & 4.6 & 4.7 & 4.15 |
| $D(s, y)$ | Distance between store $s$ and $y$ | 4.5 & 4.8 & 4.9 & 4.11 |
| $e_{i,s}$ | Price elasticity of product $i$ sold in store $x$ | 4.5 & 4.8 |
| $\psi_\cup$ | Total revenue of all products sold in a PG | 4.6 & 4.8 |
| $\psi_\cap(x, y)$ | Total revenue of all products exclusively sold in both stores $x$ and $y$ | 4.7 & 4.8 |
| $c_z$ | Cluster center $z$ with $z$ the number of the cluster | 4.9 & 4.11 |
| $\lambda$ | Potential clustering of stores | 4.12 & 4.13 & 4.14 |
| $\lambda_g$ | Cluster function of product group $g$ | 4.2 & 4.3 & 4.13 & 4.14 & 4.16 |
| $\Pi$ | Permutation matrix | 4.12 |
| $\widetilde{\lambda}$ | Cluster membership of stores | 4.12 & 4.16 |
| $\lambda^*$ | Cluster consensus, final clustering | 4.13 & 4.13 & 4.14 |
| $x_g$ | Weight of product group $g$ | 4.13 & 4.14 & 4.16 & 4.15 |
| $\Pi_g$ | Permutation matrix for PG $g$ | 4.14 & 4.16 |

**Table A.2:** The variables used in the regression analysis.

| Variable | Description |
|---|---|
| Ascension day | holiday, when holiday date falls in week then 1 else 0 |
| Back to school | school, when holiday date falls in week then 1 else 0 |
| Carnival | holiday, when holiday date falls in week then 1 else 0 |
| Easter | holiday, when holiday date falls in week then 1 else 0 |
| Father's day | holiday, when holiday date falls in week then 1 else 0 |
| Autumn | school, when holiday date falls in week then 1 else 0 |
| Krokus | school, when holiday date falls in week then 1 else 0 |
| Liberation day | holiday, when holiday date falls in week then 1 else 0 |
| New year | holiday, when holiday date falls in week then 1 else 0 |
| Pre-Easter | holiday, when holiday date falls in week then 1 else 0 |
| Pre-Pentecost | holiday, when holiday date falls in week then 1 else 0 |
| Pre-Christmas | holiday, when holiday date falls in week then 1 else 0 |
| Queen's day | holiday, when holiday date falls in week then 1 else 0 |
| Mother's day | holiday, when holiday date falls in week then 1 else 0 |
| Olympus | holiday, when holiday date falls in week then 1 else 0 |
| Pentecost | holiday, when holiday date falls in week then 1 else 0 |
| Sinterklaas | holiday, when holiday date falls in week then 1 else 0 |
| Christmas | holiday, when holiday date falls in week then 1 else 0 |
| | |
| Price ratio competition | Competitor price vs. average competitor price |
| Price ratio regular price | Regular price vs. average regular price |
| Price ratio substitute 1 | Regular price substitute 1 vs. average regular price substitute 1 |
| Price ratio substitute 2 | Regular price substitute 2 vs. average regular price substitute 2 |
| Price ratio complement 1 | Regular price complement 1 vs. average regular price complement 1 |
| | |
| Discount price | Discount amount on regular price |
| Trips | Number of transactions |
| Seasonality | Captures the overall sales trend |
| Rain | Total mm rain in a week |
| Temperature | Average temperature in Kelvin in a week |
| | |
| Instore flyer | Promotion dummy |
| Signing on shelf | Promotion dummy |
| Openings flyer | Promotion dummy |
| Dynamic newsletter | Promotion dummy |
| Entrance poster | Promotion dummy |
| National folder | Promotion dummy |
| Buy-1-get-1-free | Promotion dummy |
| Second placing | Promotion dummy |
| Online radio | Promotion dummy |
| Store radio | Promotion dummy |
| | |
| Average family size | Number of persons versus the number of household |
| City-rural dummy | When city then 1 else 0 |
| Franchise dummy | Own store indicator |
| Holiday indicator (e.g. Christmas) | When holiday date falls in week then 1 else 0 |
| No. competitive stores in 15km | Number of competitors in a radius of 15km from a store |
| Percentage high income | Share of consumer with high income (above 1.5x modal) |
| Percentage high social class | Share of of consumers part of the high social class |
| Percentage high education | Share of high educated consumers |
| Percentage low education | Share of low educated consumers |
| Percentage regular visitor competitor | Share of consumers that regularly visits the biggest competitor |
| Percentage unemployed | Share of unemployed consumers |
| Percentage western immigrants | Share of western immigrants |
| Percentage non-western immigrants | Share of non-western immigrants |
| Region East | Stores situated in the East of the country |
| Region South | Stores situated in the South of the country |
| Region West | Stores situated in the West of the country |
| Self-scan dummy | When self-scanner available then 1 else 0 |
| Surface in sqm | Surface of supermarket in square meters |

**Table A.3:** The social class, the division is based upon the breadwinner of a family. Horizontally we can see the study degree according the Dutch education system and vertically the profession.

| **Division social classes** **Profession breadwinner** | WO | HBO | HAVO/VWO | MBO | MAVO | LBO/VMBO | LO and not specified |
|---|---|---|---|---|---|---|---|
| Business owner (leads 10 or more persons) | A (high) | A | A | A | B1 | B2 | C |
| Business owner (leads less than 10 persons) | A | A | A | A | B1 | B2 | C |
| Farmer / gardener | A | A | B1 | B1 | B2 | C | C |
| Free professions | A | A | A | B1 | B2 | C | C |
| Scientist and higher; manager | A | B1 | B1 | B1 | B2 | C | C |
| Scientist and higher; no manager | A | B1 | B1 | B1 | B2 | C | C |
| Secondary technical and vocational education (SBC 4); manager | A | B1 | B1 | B1 | B2 | C | C |
| Secondary technical and vocational education (SBC 4); no manager | A | B1 | B1 | B2 | B2 | C | C |
| Elementary school and lower technical and vocational education (SBC 1 en 2) | B1 | B2 | B2 | C | C | C | C |
| Early retirement (dutch: VUT) / vocational Retirement | A | A | B1 | B1 | B2 | C | D |
| Unemployed / Disabled / Social welfare provision | B1 | B2 | C | C | C | C | D |
| Student / Other | B1 | B2 | C | C | D | D | D (low) |

# B | Impact clustering results

**Table B.1:** The step wise revenue without taking competition into account, the horizontal percentage change represents the price increase in inelastic products and the vertical percentage change represents the price cut of elastic products (in thousands of euros).

| | +0% | +1% | +2% | +3% | +4% | +5% |
|---|---|---|---|---|---|---|
| -0% | 148,406 | 148,491 | 148,574 | 148,656 | 148,737 | 148,817 |
| -1% | 148,537 | 148,622 | 148,706 | 148,788 | 148,869 | 148,949 |
| -2% | 148,664 | 148,748 | 148,832 | 148,914 | 148,995 | 149,075 |
| -3% | 148,785 | 148,869 | 148,953 | 149,035 | 149,116 | 149,196 |
| -4% | 148,900 | 148,985 | 149,069 | 149,151 | 149,232 | 149,312 |
| -5% | 149,011 | 149,096 | 149,179 | 149,262 | 149,343 | *149,422* |

**Table B.2:** The step wise profit without taking competition into account, the horizontal percentage change represents the price increase in inelastic products and the vertical percentage change represents the price cut of elastic products (in thousands of euros).

| | +0% | +1% | +2% | +3% | +4% | +5% |
|---|---|---|---|---|---|---|
| -0% | 39,661 | 39,792 | 39,921 | 40,049 | 40,175 | *40,301* |
| -1% | 39,597 | 39,727 | 39,856 | 39,984 | 40,111 | 40,236 |
| -2% | 39,527 | 39,657 | 39,787 | 39,914 | 40,041 | 40,166 |
| -3% | 39,452 | 39,582 | 39,712 | 39,839 | 39,966 | 40,091 |
| -4% | 39,372 | 39,502 | 39,631 | 39,759 | 39,886 | 40,011 |
| -5% | 39,287 | 39,417 | 39,546 | 39,674 | 39,801 | 39,926 |

**Table B.3:** The projected increase in profit with competition vs. without competition taken into account and 2% price change of elastic products, between brackets the relative change can be found.

| Product group | Competition | No competition |
|---|---|---|
| Cola | 62,687 (0.63%) | 171,021 (1.71%) |
| Energy | 13,134 (0.31%) | 15,301 (0.37%) |
| Juice | 56,333 (0.76%) | 65,075 (0.88%) |
| Kids youth | 65,890 (1.10%) | 85,699 (1.43%) |
| Large soda | 76,465 (0.99%) | 126,722 (1.65%) |
| Water | 25,683 (0.59%) | 41,388 (0.95%) |
| Total | 300,193 (0.76%) | 505,207 (1.26%) |

**Table B.4:** The share of products labeled as inelastic or elastic by the labeling procedure (in percentages).

| Clusters | Cola | | Energy | | Juices | | Kids youth | | Large soda | | Water | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Inelas. | Elas. | Inelas. | Elas. | Inelas. | Elas. | Inelas. | Elas. | Inelas. | Elas. | Inelas. | Elas. |
| Cluster 1 | 10% | 13% | 0% | 2% | 3% | 17% | 7% | 11% | 5% | 18% | 10% | 14% |
| Cluster 2 | 5% | 3% | 6% | 18% | 3% | 2% | 4% | 3% | 21% | 7% | 3% | 3% |
| Cluster 3 | 23% | 8% | 12% | 11% | 7% | 10% | 4% | 16% | 5% | 2% | 8% | 9% |
| Cluster 4 | 9% | 11% | 3% | 9% | 17% | 14% | 13% | 5% | 7% | 8% | %11 | 10% |
| Total | 11% | 8% | 4% | 8% | 7% | 9% | 8% | 7% | 9% | 7% | 7% | 8% |

# C | Panel data regression results

**Table C.1:** The table presents the coefficients (and standard errors) of the variables of the panel data regression models for Cola on store-item level for all clusters to declare the dependent variable of price elasticities. Note, that the significance level is set to 5%.

| **Panel data regression model**<br>**Independent variables** | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| Intercept | 1.120 (0.050) | -1.755 (0.059) | -0.011 (0.062) | -0.127 (0.100) |
| | | | | |
| Ascension day | 0.022 (0.005) | 0.044 (0.006) | 0.009 (0.006) | 0.026 (0.009) |
| Back to school | 0.004 (0.004) | 0.017 (0.005) | 0.008 (0.005) | -0.004 (0.007) |
| Carnival | 0.025 (0.005) | 0.000 (0.006) | 0.005 (0.006) | 0.023 (0.009) |
| Easter | 0.007 (0.004) | 0.024 (0.005) | 0.000 (0.005) | 0.011 (0.007) |
| Father's day | -0.025 (0.004) | -0.033 (0.005) | -0.027 (0.005) | -0.027 (0.008) |
| Autumn | 0.013 (0.004) | 0.031 (0.005) | 0.027 (0.005) | 0.017 (0.007) |
| Krokus | -0.001 (0.004) | -0.024 (0.005) | 0.004 (0.005) | -0.012 (0.007) |
| Liberation day | -0.005 (0.004) | -0.020 (0.005) | -0.012 (0.005) | -0.001 (0.008) |
| New year | 0.045 (0.004) | 0.021 (0.005) | 0.035 (0.005) | 0.055 (0.007) |
| Pre-Easter | -0.019 (0.004) | 0.004 (0.005) | -0.018 (0.005) | -0.013 (0.007) |
| Pre-Pentecost | -0.020 (0.004) | -0.016 (0.005) | -0.022 (0.005) | -0.024 (0.008) |
| Pre-Christmas | 0.033 (0.007) | 0.053 (0.010) | 0.034 (0.010) | 0.052 (0.014) |
| Queen's day | -0.011 (0.004) | -0.005 (0.005) | -0.013 (0.005) | 0.003 (0.008) |
| Mother's day | 0.050 (0.006) | 0.054 (0.007) | 0.029 (0.007) | 0.069 (0.012) |
| Olympus | 0.111 (0.007) | 0.086 (0.008) | 0.110 (0.008) | 0.082 (0.012) |
| Pentecost | -0.001 (0.004) | 0.009 (0.005) | 0.007 (0.005) | 0.000 (0.008) |
| Sinterklaas | 0.046 (0.004) | 0.062 (0.005) | 0.027 (0.005) | 0.066 (0.007) |
| Christmas | 0.048 (0.006) | 0.070 (0.008) | 0.057 (0.008) | 0.083 (0.010) |
| | | | | |
| Price ratio complement 1 | 0.010 (0.017) | 0.031 (0.020) | -0.053 (0.020) | -0.091 (0.032) |
| Price ratio competition | -0.636 (0.049) | 0.348 (0.064) | -0.224 (0.066) | -0.179 (0.096) |
| Price ratio regular price | 0.145 (0.062) | 1.773 (0.078) | 1.004 (0.082) | 0.912 (0.121) |
| Price ratio substitute 1 | -0.884 (0.015) | -0.591 (0.018) | -0.834 (0.019) | -0.662 (0.024) |
| Price ratio substitute 2 | -0.780 (0.032) | -1.202 (0.038) | -0.921 (0.044) | -1.127 (0.069) |
| | | | | |
| Discount price | -0.005 (0.000) | -0.006 (0.000) | -0.005 (0.000) | -0.004 (0.001) |
| Trips | 0.006 (0.000) | 0.003 (0.000) | 0.005 (0.000) | 0.012 (0.000) |
| | | | | |
| Dynamic newsletter | 0.000 (0.000) | 0.643 (0.106) | 0.463 (0.111) | 1.366 (0.153) |
| Entrance poster | -1.647 (0.502) | -5.129 (0.593) | -1.856 (0.586) | 0.000 (0.000) |
| National folder | 0.825 (0.114) | 0.704 (0.145) | 0.765 (0.162) | 0.876 (0.229) |
| Instore flyer | -0.348 (0.042) | -0.806 (0.091) | -0.602 (0.095) | -1.428 (0.133) |
| Online radio | 0.134 (0.029) | -0.116 (0.035) | 0.162 (0.039) | 0.237 (0.057) |
| Openings flyer | -1.979 (0.124) | -2.005 (0.148) | -0.875 (0.143) | -1.356 (0.138) |
| Signing on shelf | 1.341 (0.064) | 0.362 (0.079) | 1.042 (0.081) | 0.759 (0.125) |
| Second placing | -0.814 (0.071) | 0.471 (0.089) | -0.239 (0.095) | 0.620 (0.138) |
| Store radio | 3.76 (0.486) | 6.577 (0.580) | 2.586 (0.573) | 0.000 (0.000) |
| | | | | |
| Rain (x1000) | 0.017 (0.003) | 0.015 (0.004) | 0.012 (0.005 ) | 0.035 (0.007) |
| Temperature (x1000) | 0.221 (0.019) | 0.617 (0.016) | 0.213 (0.016) | 0.644 (0.023) |
| Seasonality | 0.098 (0.006) | 0.155 (0.009) | 0.130 (0.007) | 0.000 (0.000) |
| | | | | |
| Percentage high education | -0.306 (0.069) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| Percentage high income | 0.158 (0.05) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| City-rural dummy | -0.035 (0.012) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| No. competitive stores in 15km | 0.000 (0.000) | -0.013 (0.003) | -0.005 (0.002) | 0.000 (0.000) |
| $R^2$ | 0.575 | 0.443 | 0.562 | 0.625 |

**Table C.2:** The table presents the coefficients (and standard errors) of the variables of the panel data regression models for Energy on store-item level for all clusters to declare the dependent variable of price elasticities. Note, that the significance level is set to 5%.

| **Panel data regression model**<br>**Independent variables** | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| Intercept | 1.562 (0.061) | 0.315 (0.056) | 1.045 (0.086) | 0.957 (0.087) |
| | | | | |
| Ascension day | 0.076 (0.007) | 0.064 (0.005) | 0.060 (0.009) | 0.049 (0.010) |
| Back to school | -0.001 (0.006) | -0.014 (0.004) | 0.017 (0.009) | 0.005 (0.009) |
| Carnival | 0.013 (0.008) | 0.015 (0.006) | 0.022 (0.011) | 0.007 (0.012) |
| Easter | 0.045 (0.006) | 0.018 (0.004) | 0.031 (0.008) | 0.022 (0.009) |
| Father's day | 0.017 (0.007) | 0.012 (0.004) | 0.019 (0.009) | 0.001 (0.010) |
| Autumn | -0.063 (0.007) | -0.023 (0.005) | -0.070 (0.009) | -0.025 (0.009) |
| Krokus | 0.023 (0.006) | 0.015 (0.004) | 0.024 (0.008) | 0.017 (0.009) |
| Liberation day | 0.017 (0.007) | -0.006 (0.005) | 0.016 (0.009) | -0.004 (0.010) |
| New year | -0.060 (0.006) | -0.030 (0.004) | -0.084 (0.008) | -0.057 (0.009) |
| Pre-Easter | 0.045 (0.006) | 0.021 (0.004) | 0.050 (0.008) | 0.029 (0.009) |
| Pre-Pentecost | 0.029 (0.007) | 0.026 (0.005) | 0.026 (0.010) | 0.020 (0.010) |
| Pre-Christmas | -0.010 (0.012) | -0.013 (0.009) | -0.008 (0.017) | -0.029 (0.018) |
| Queen's day | 0.073 (0.007) | 0.037 (0.005) | 0.074 (0.009) | 0.033 (0.010) |
| Mother's day | 0.105 (0.011) | 0.038 (0.007) | 0.092 (0.013) | 0.066 (0.016) |
| Olympus | 0.253 (0.013) | 0.099 (0.008) | 0.239 (0.015) | 0.173 (0.024) |
| Pentecost | 0.053 (0.007) | 0.041 (0.004) | 0.041 (0.009) | 0.019 (0.010) |
| Sinterklaas | -0.040 (0.006) | -0.019 (0.004) | -0.033 (0.008) | -0.038 (0.009) |
| Christmas | -0.012 (0.009) | 0.005 (0.007) | -0.009 (0.012) | 0.016 (0.013) |
| | | | | |
| Price ratio competition | -1.464 (0.045) | -0.541 (0.031) | -1.263 (0.061) | -0.531 (0.053) |
| Price ratio regular price | -0.390 (0.071) | -0.550 (0.051) | -0.047 (0.100) | -1.404 (0.106) |
| Price ratio substitute 1 | -0.361 (0.029) | -0.362 (0.020) | -0.158 (0.039) | -0.319 (0.040) |
| Price ratio substitute 2 | -0.564 (0.025) | -0.330 (0.016) | -0.742 (0.032) | -0.337 (0.035) |
| | | | | |
| Discount price | 0.016 (0.003) | 0.015 (0.002) | 0.016 (0.006) | 0.033 (0.006) |
| Trips | 0.001 (0.000) | 0.004 (0.000) | -0.008 (0.000) | 0.001 (0.000) |
| | | | | |
| Rain (x1000) | -0.083 (0.006) | -0.032 (0.004) | -0.101 (0.008) | -0.084 (0.009) |
| Temperature (x1000) | 0.620 (0.020) | 0.608 (0.013) | 0.690 (0.030) | 0.620 (0.028) |
| Seasonality | 0.232 (0.01) | 0.178 (0.004) | 0.236 (0.014) | 0.252 (0.006) |
| | | | | |
| National folder | -6.656 (0.391) | -4.993 (0.255) | -6.690 (0.502) | -6.671 (0.502) |
| Instore flyer | -0.795 (0.147) | -0.757 (0.096) | -1.231 (0.186) | -1.203 (0.196) |
| Online radio | 0.221 (0.035) | 0.084 (0.022) | 0.286 (0.049) | 3.332 (0.173) |
| Signing on shelf | 2.612 (0.138) | 2.328 (0.089) | 2.585 (0.172) | 0.776 (0.076) |
| Second placing | 0.480 (0.053) | 0.639 (0.035) | 0.401 (0.075) | 0.000 (0.000) |
| | | | | |
| Region East | 0.000 (0.000) | 0.115 (0.055) | 0.000 (0.000) | 0.000 (0.000) |
| Region South | 0.000 (0.000) | 0.139 (0.042) | 0.000 (0.000) | 0.000 (0.000) |
| City-rural dummy | 0.000 (0.000) | -0.061 (0.019) | 0.000 (0.000) | 0.000 (0.000) |
| Percentage high education | 0.000 (0.000) | 0.000 (0.000) | 0.500 (0.124) | 0.334 (0.018) |
| No. competitive stores in 15km | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.002 (0.001) |
| $R^2$ | 0.576 | 0.683 | 0.555 | 0.657 |

**Table C.3:** The table presents the coefficients (and standard errors) of the variables of the panel data regression models for Kids youth on store-item level for all clusters to declare the dependent variable of price elasticities. Note, that the significance level is set to 5%.

| **Panel data regression model<br>Independent variables** | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| Intercept | -0.637 (0.059) | -0.393 (0.056) | -0.808 (0.078) | -1.000 (0.084) |
| Ascension day | -0.004 (0.006) | 0.026 (0.005) | 0.002 (0.007) | 0.004 (0.010) |
| Back to school | 0.001 (0.006) | 0.042 (0.005) | -0.028 (0.007) | -0.008 (0.010) |
| Carnival | -0.021 (0.007) | -0.027 (0.006) | -0.023 (0.009) | -0.017 (0.012) |
| Easter | -0.046 (0.006) | -0.056 (0.005) | -0.062 (0.007) | -0.050 (0.009) |
| Father's day | -0.086 (0.006) | -0.039 (0.005) | -0.098 (0.007) | -0.079 (0.010) |
| Autumn | -0.015 (0.006) | -0.025 (0.005) | -0.026 (0.007) | -0.017 (0.009) |
| Krokus | 0.028 (0.006) | 0.022 (0.005) | 0.045 (0.007) | 0.036 (0.009) |
| Liberation day | -0.001 (0.006) | -0.013 (0.005) | -0.021 (0.007) | -0.024 (0.010) |
| New year | -0.021 (0.006) | -0.082 (0.006) | -0.005 (0.008) | -0.058 (0.010) |
| Pre-Easter | -0.061 (0.005) | -0.045 (0.005) | -0.076 (0.007) | -0.073 (0.009) |
| Pre-Pentecost | -0.003 (0.006) | 0.013 (0.005) | -0.001 (0.008) | -0.004 (0.010) |
| Pre-Christmas | 0.087 (0.011) | 0.053 (0.010) | 0.061 (0.014) | 0.071 (0.018) |
| Queen's day | -0.009 (0.006) | -0.015 (0.006) | -0.031 (0.008) | -0.023 (0.011) |
| Mother's day | -0.010 (0.008) | -0.028 (0.007) | 0.006 (0.010) | -0.032 (0.014) |
| Olympus | 0.078 (0.008) | 0.031 (0.007) | 0.097 (0.009) | 0.063 (0.014) |
| Pentecost | 0.054 (0.006) | 0.036 (0.005) | 0.041 (0.007) | 0.043 (0.010) |
| Sinterklaas | -0.022 (0.006) | -0.036 (0.005) | -0.012 (0.007) | -0.018 (0.009) |
| Christmas | -0.029 (0.008) | -0.019 (0.008) | -0.015 (0.011) | -0.035 (0.013) |
| Price ratio competition | -0.379 (0.096) | -0.554 (0.088) | 0.176 (0.121) | -0.566 (0.163) |
| Price ratio regular price | 0.420 (0.089) | 0.267 (0.080) | -0.199 (0.113) | 0.676 (0.152) |
| Price ratio substitute 1 | 0.601 (0.050) | -0.404 (0.041) | 0.251 (0.069) | 0.413 (0.079) |
| Price ratio substitute 2 | -1.024 (0.048) | 0.115 (0.039) | -0.715 (0.062) | -0.769 (0.077) |
| Discount price | 0.002 (0.004) | 0.031 (0.003) | 0.013 (0.005) | 0.003 (0.009) |
| Trips | 0.027 (0.001) | 0.007 (0.000) | 0.025 (0.001) | 0.022 (0.001) |
| National folder | -0.499 (0.207) | -0.351 (0.181) | -1.579 (0.260) | -0.093 (0.330) |
| Instore flyer | 1.598 (0.132) | 0.681 (0.113) | 1.494 (0.169) | 1.288 (0.218) |
| Online radio | 0.087 (0.038) | 0.619 (0.032) | 0.073 (0.049) | 0.264 (0.062) |
| Signing on shelf | -2.588 (0.139) | -1.661 (0.121) | -2.626 (0.175) | -2.033 (0.222) |
| Second placing | 0.087 (0.113) | 0.769 (0.109) | 0.724 (0.147) | -0.025 (0.178) |
| Openings flyer | 2.362 (0.078) | 1.205 (0.072) | 1.774 (0.102) | 2.169 (0.122) |
| Rain (x1000) | 0.014 (0.002) | -0.037 (0.005) | 0.000 (0.000) | 0.021 (0.009) |
| Temperature (x1000) | 0.325 (0.019) | 0.260 (0.018) | 0.351 (0.023) | 0.500 (0.030) |
| Seasonality | 0.034 (0.008) | 0.110 (0.009) | 0.000 (0.000) | 0.082 (0.007) |
| No. competitive store in 15km | 0.007 (0.002) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| Percentage unemployed | -1.619 (0.333) | -1.721 (0.373) | 0.000 (0.000) | 0.000 (0.000) |
| Percentage regular visitor competitor | -0.137 (0.039) | -0.143 (0.048) | 0.156 (0.061) | 0.086 (0.025) |
| Percentage high education | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.199 (0.033) |
| Percentage high income | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | -0.144 (0.029) |
| $R^2$ | 0.530 | 0.328 | 0.520 | 0.602 |

**Table C.4:** The table presents the coefficients (and standard errors) of the variables of the panel data regression models for Large soda on store-item level for all clusters to declare the dependent variable of price elasticities. Note, that the significance level is set to 5%.

| **Panel data regression model**<br>**Independent variables** | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| Intercept | 2.114 (0.055) | 0.887 (0.050) | 2.524 (0.07) | 1.918 (0.093) |
| Ascension day | 0.043 (0.006) | 0.039 (0.005) | 0.035 (0.008) | 0.036 (0.011) |
| Back to school | -0.020 (0.005) | -0.019 (0.005) | -0.023 (0.007) | -0.033 (0.010) |
| Carnival | 0.071 (0.007) | 0.036 (0.006) | 0.113 (0.009) | 0.045 (0.013) |
| Easter | 0.055 (0.005) | 0.053 (0.005) | 0.043 (0.007) | 0.036 (0.009) |
| Father's day | -0.001 (0.006) | -0.009 (0.005) | 0.001 (0.008) | -0.016 (0.011) |
| Autumn | -0.043 (0.005) | -0.043 (0.005) | -0.031 (0.007) | -0.053 (0.010) |
| Krokus | 0.011 (0.005) | 0.017 (0.005) | -0.009 (0.007) | 0.001 (0.010) |
| Liberation day | 0.031 (0.006) | 0.013 (0.005) | 0.030 (0.008) | 0.013 (0.011) |
| New year | 0.057 (0.005) | 0.066 (0.005) | 0.016 (0.007) | 0.075 (0.010) |
| Pre-Easter | 0.092 (0.005) | 0.162 (0.005) | 0.027 (0.007) | 0.101 (0.009) |
| Pre-Pentecost | 0.000 (0.006) | 0.026 (0.005) | -0.015 (0.008) | 0.006 (0.011) |
| Pre-Christmas | 0.066 (0.011) | 0.072 (0.010) | 0.040 (0.014) | 0.096 (0.019) |
| Queen's day | -0.004 (0.006) | 0.02 (0.005) | -0.011 (0.008) | -0.016 (0.011) |
| Mother's day | 0.004 (0.009) | -0.002 (0.007) | 0.007 (0.011) | 0.018 (0.017) |
| Olympus | 0.141 (0.007) | 0.086 (0.006) | 0.128 (0.009) | 0.147 (0.015) |
| Pentecost | 0.069 (0.006) | 0.055 (0.005) | 0.060 (0.007) | 0.056 (0.010) |
| Sinterklaas | -0.070 (0.005) | -0.046 (0.005) | -0.096 (0.007) | -0.059 (0.010) |
| Christmas | 0.031 (0.008) | 0.030 (0.008) | 0.028 (0.011) | 0.056 (0.014) |
| Price ratio competition | -0.343 (0.091) | -0.221 (0.080) | -0.892 (0.130) | -0.189 (0.175) |
| Price ratio regular price | -3.586 (0.114) | -2.459 (0.100) | -3.226 (0.168) | -3.386 (0.226) |
| Price ratio substitute 1 | -0.282 (0.059) | -0.266 (0.048) | -0.413 (0.080) | -0.267 (0.105) |
| Price ratio substitute 2 | 0.308 (0.039) | 0.213 (0.031) | 0.403 (0.053) | 0.041 (0.065) |
| Price index complement 1 | 0.601 (0.043) | 0.474 (0.035) | 0.549 (0.057) | 0.472 (0.078) |
| Discount price | 0.071 (0.002) | 0.042 (0.001) | 0.038 (0.002) | 0.102 (0.006) |
| Trips | 0.014 (0.001) | 0.019 (0.000) | 0.013 (0.001) | 0.009 (0.001) |
| National folder | -0.820 (0.203) | -0.870 (0.168) | -0.639 (0.283) | -0.621 (0.392) |
| Instore flyer | 0.812 (0.184) | 0.888 (0.153) | 0.842 (0.259) | 0.605 (0.353) |
| Online radio | 0.011 (0.028) | -0.151 (0.023) | 0.269 (0.036) | -0.116 (0.051) |
| Signing on shelf | -1.036 (0.246) | -1.206 (0.205) | -1.412 (0.354) | 0.084 (0.461) |
| Second placing | 1.106 (0.143) | 1.201 (0.121) | 1.044 (0.208) | 0.363 (0.264) |
| Openings flyer | -0.154 (0.091) | 0.279 (0.077) | -0.135 (0.127) | 0.000 (0.000) |
| Rain (x1000) | -0.049 (0.005) | -0.040 (0.005) | -0.057 (0.007) | -0.040 (0.010) |
| Temperature (x1000) | -0.129 (0.020) | -0.037 (0.018) | -0.242 (0.028) | 0.000 (0.000) |
| Seasonality | 0.170 (0.007) | 0.203 (0.007) | 0.122 (0.010) | 0.145 (0.010) |
| No. competitive stores in 15km | -0.006 (0.002) | -0.005 (0.002) | 0.000 (0.000) | 0.000 (0.000) |
| Region East | -0.029 (0.013) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| Region South | -0.072 (0.017) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| City-rural dummy | -0.051 (0.012) | 0.000 (0.000) | -0.074 (0.027) | 0.000 (0.000) |
| Self-scan dummy | -0.041 (0.012) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| Percentage unemployed | 0.000 (0.000) | -1.345 (0.413) | 0.000 (0.000) | 0.000 (0.000) |
| Percentage non-western immigrants | 0.000 (0.000) | 0.567 (0.189) | 0.000 (0.000) | 0.000 (0.000) |
| Percentage western-immigrants | 0.000 (0.000) | -0.304 (0.146) | 0.000 (0.000) | 0.000 (0.000) |
| $R^2$ | 0.550 | 0.580 | 0.543 | 0.612 |

**Table C.5:** The table presents the coefficients (and standard errors) of the variables of the panel data regression models for Juices on store-item level for all clusters to declare the dependent variable of price elasticities. Note, that the significance level is set to 5%.

| **Panel data regression model**<br>**Independent variables** | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| Intercept | -4.936 (0.070) | -5.774 (0.086) | -5.009 (0.106) | -4.848 (0.121) |
| | | | | |
| Ascension day | 0.125 (0.010) | 0.111 (0.010) | 0.122 (0.012) | 0.132 (0.015) |
| Back to school | -0.047 (0.009) | -0.042 (0.009) | -0.074 (0.012) | -0.046 (0.014) |
| Carnival | -0.045 (0.011) | -0.042 (0.011) | -0.043 (0.014) | -0.046 (0.017) |
| Easter | 0.099 (0.009) | 0.125 (0.009) | 0.098 (0.011) | 0.099 (0.012) |
| Father's day | 0.090 (0.010) | 0.090 (0.009) | 0.070 (0.012) | 0.079 (0.014) |
| Autumn | 0.056 (0.009) | 0.078 (0.009) | 0.050 (0.011) | 0.032 (0.013) |
| Krokus | 0.016 (0.009) | -0.017 (0.009) | 0.055 (0.011) | 0.028 (0.013) |
| Liberation day | 0.012 (0.010) | 0.010 (0.009) | 0.013 (0.012) | 0.013 (0.014) |
| New year | 0.006 (0.009) | -0.012 (0.009) | 0.015 (0.012) | 0.016 (0.013) |
| Pre-Easter | 0.111 (0.009) | 0.182 (0.009) | 0.045 (0.011) | 0.136 (0.013) |
| Pre-Pentecost | 0.107 (0.010) | 0.136 (0.010) | 0.059 (0.012) | 0.105 (0.015) |
| Pre-Christmas | 0.072 (0.018) | 0.059 (0.018) | 0.111 (0.023) | 0.138 (0.027) |
| Queen's day | 0.022 (0.009) | 0.029 (0.009) | 0.025 (0.012) | 0.017 (0.014) |
| Mother's day | -0.001 (0.014) | -0.023 (0.012) | 0.028 (0.016) | -0.002 (0.021) |
| Olympus | -0.159 (0.013) | -0.177 (0.011) | -0.069 (0.015) | -0.170 (0.020) |
| Pentecost | 0.028 (0.009) | 0.018 (0.009) | 0.032 (0.012) | 0.039 (0.014) |
| Sinterklaas | 0.019 (0.009) | -0.006 (0.009) | 0.081 (0.011) | 0.074 (0.013) |
| Christmas | 0.132 (0.013) | 0.155 (0.014) | 0.129 (0.018) | 0.094 (0.019) |
| | | | | |
| Price ratio competition | 0.807 (0.104) | 1.017 (0.098) | 0.192 (0.132) | 0.644 (0.164) |
| Price ratio regular price | 3.308 (0.138) | 3.184 (0.129) | 4.186 (0.178) | 2.991 (0.206) |
| Price ratio substitute 1 | 1.773 (0.045) | 1.859 (0.040) | 1.224 (0.054) | 1.700 (0.072) |
| Price ratio substitute 2 | -2.984 (0.045) | -2.81 (0.040) | -2.868 (0.057) | -2.746 (0.066) |
| | | | | |
| Discount price | 0.129 (0.013) | 0.045 (0.006) | 0.021 (0.005) | 0.122 (0.015) |
| Trips | 0.024 (0.001) | 0.017 (0.001) | 0.030 (0.001) | 0.016 (0.001) |
| | | | | |
| Instore flyer | 0.945 (0.193) | 0.584 (0.169) | 1.425 (0.24) | 0.655 (0.291) |
| Online radio | -0.587 (0.047) | -0.319 (0.044) | -0.842 (0.058) | -0.537 (0.071) |
| Signing on shelf | 2.286 (0.365) | 2.027 (0.333) | 2.502 (0.476) | 2.463 (0.537) |
| Second placing | -0.006 (0.317) | 1.209 (0.291) | -0.101 (0.418) | 0.092 (0.473) |
| Openings flyer | -0.612 (0.223) | -1.829 (0.212) | -0.535 (0.293) | -1.136 (0.310) |
| | | | | |
| Rain (x1000) | 0.064 (0.008) | 0.090 (0.009) | 0.051 (0.011) | 0.089 (0.013) |
| Temperature (x1000) | -0.123 (0.028) | 0.000 (0.000) | -0.225 (0.037) | 0.000 (0.000) |
| Seasonality | 0.285 (0.015) | 0.343 (0.015) | 0.319 (0.019) | 0.306 (0.014) |
| | | | | |
| Franchise dummy | -0.052 (0.022) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| Percentage high education | 0.000 (0.000) | 0.195 (0.077) | 0.000 (0.000) | 0.000 (0.000) |
| Region East | 0.000 (0.000) | 0.188 (0.061) | 0.000 (0.000) | 0.000 (0.000) |
| Region South | 0.000 (0.000) | 0.285 (0.049) | 0.000 (0.000) | 0.000 (0.000) |
| Percentage high social class | 0.000 (0.000) | 0.000 (0.000) | 0.312 (0.130) | 0.000 (0.000) |
| Percentage regular visitor competitor | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.312 (0.096) |
| $R^2$ | 0.563 | 0.531 | 0.613 | 0.589 |

**Table C.6:** The table presents the coefficients (and standard errors) of the variables of the panel data regression models for Water on store-item level for all clusters to declare the dependent variable of price elasticities. Note, that the significance level is set to 5%.

| **Panel data regression model**<br>**Independent variables** | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| Intercept | -0.938 (0.030) | -1.585 (0.036) | -0.782 (0.036) | -1.097 (0.059) |
| | | | | |
| Ascension day | 0.026 (0.004) | 0.020 (0.005) | 0.040 (0.005) | 0.019 (0.008) |
| Back to school | -0.005 (0.004) | -0.020 (0.004) | -0.011 (0.005) | 0.006 (0.008) |
| Carnival | -0.006 (0.005) | -0.032 (0.006) | -0.002 (0.006) | -0.009 (0.010) |
| Easter | 0.015 (0.004) | 0.015 (0.004) | 0.011 (0.005) | 0.023 (0.007) |
| Father's day | -0.007 (0.004) | -0.006 (0.004) | -0.004 (0.005) | -0.011 (0.008) |
| Autumn | -0.009 (0.004) | -0.001 (0.004) | 0.009 (0.005) | -0.008 (0.008) |
| Krokus | 0.014 (0.004) | -0.003 (0.004) | 0.012 (0.005) | 0.009 (0.008) |
| Liberation day | 0.006 (0.004) | 0.016 (0.005) | -0.023 (0.005) | -0.002 (0.008) |
| New year | -0.025 (0.004) | -0.027 (0.005) | -0.041 (0.005) | -0.030 (0.007) |
| Pre-Easter | 0.018 (0.004) | 0.021 (0.004) | 0.013 (0.005) | 0.019 (0.007) |
| Pre-Pentecost | 0.027 (0.004) | 0.041 (0.005) | 0.012 (0.005) | 0.032 (0.009) |
| Pre-Christmas | 0.058 (0.008) | 0.073 (0.009) | 0.093 (0.010) | 0.056 (0.015) |
| Queen's day | 0.018 (0.004) | 0.016 (0.005) | 0.031 (0.005) | 0.001 (0.008) |
| Mother's day | -0.017 (0.006) | -0.039 (0.006) | -0.005 (0.007) | -0.020 (0.012) |
| Olympus | -0.001 (0.007) | -0.013 (0.008) | 0.021 (0.009) | 0.042 (0.016) |
| Pentecost | 0.023 (0.004) | 0.025 (0.004) | 0.029 (0.005) | 0.028 (0.008) |
| Sinterklaas | 0.012 (0.004) | 0.045 (0.004) | 0.001 (0.005) | 0.017 (0.007) |
| Christmas | 0.045 (0.006) | 0.066 (0.007) | 0.031 (0.007) | 0.046 (0.011) |
| | | | | |
| Price ratio competition | -0.065 (0.032) | -0.218 (0.039) | -0.105 (0.038) | -0.320 (0.060) |
| Price ratio regular price | -0.089 (0.041) | 0.563 (0.048) | -0.363 (0.050) | 0.339 (0.081) |
| Price ratio substitute 1 | 0.079 (0.037) | -0.310 (0.04) | 0.508 (0.044) | -0.530 (0.077) |
| Price ratio substitute 2 | -0.365 (0.024) | 0.086 (0.027) | -0.770 (0.030) | 0.187 (0.052) |
| | | | | |
| Discount price | -0.015 (0.006) | 0.005 (0.003) | 0.019 (0.004) | -0.020 (0.009) |
| Trips | 0.006 (0.000) | 0.008 (0.000) | 0.010 (0.000) | 0.000 (0.000) |
| | | | | |
| Instore flyer | 2.570 (0.445) | 3.033 (0.460) | 3.045 (0.475) | 2.084 (0.819) |
| Online radio | -0.135 (0.018) | -0.158 (0.020) | 0.047 (0.023) | -0.050 (0.035) |
| Signing on shelf | -0.177 (0.027) | -0.303 (0.03) | 0.095 (0.033) | -0.174 (0.057) |
| Second placing | -0.172 (0.392) | 0.762 (0.425) | -0.398 (0.504) | 0.773 (0.821) |
| Openings flyer | -0.444 (0.388) | -1.578 (0.421) | -0.726 (0.494) | -0.864 (0.811) |
| National folder | -1.504 (0.434) | -1.773 (0.446) | -1.706 (0.462) | -1.631 (0.792) |
| | | | | |
| Rain (x1000) | -0.028 (0.004) | -0.019 (0.004) | -0.053 (0.005) | -0.036 (0.008) |
| Temperature (x1000) | 0.625 (0.015) | 0.583 (0.019) | 0.483 (0.020) | 0.909 (0.027) |
| Seasonality | 0.224 (0.004) | 0.234 (0.005) | 0.238 (0.005) | 0.229 (0.006) |
| | | | | |
| No. competitive stores in 15km | -0.007 (0.001) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| City-rural dummy | 0.054 (0.011) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| Franchise dummy | 0.000 (0.000) | 0.000 (0.000) | -0.074 (0.028) | 0.000 (0.000) |
| Self-scan dummy | 0.000 (0.000) | 0.000 (0.000) | 0.061 (0.019) | 0.000 (0.000) |
| $R^2$ | 0.684 | 0.609 | 0.763 | 0.828 |

**Table C.7:** The table presents the coefficients (and standard errors) of the variables of the panel data regression models on store-item level for all product groups to declare the dependent variable of price elasticities. Note, that the significance level is set to 5%.

| **Panel data regression model** **Independent variables** | Cola | Energy | Large soda | Kids youth | Juices | Water |
|---|---|---|---|---|---|---|
| Intercept | 0.031 (0.054) | 0.847 (0.097) | 1.873 (0.037) | -0.418 (0.082) | -5.100 (0.044) | -1.183 (0.024) |
| Ascension day | 0.030 (0.003) | 0.064 (0.004) | 0.038 (0.003) | 0.011 (0.003) | 0.120 (0.006) | 0.026 (0.003) |
| Back to school | 0.009 (0.003) | -0.003 (0.003) | -0.024 (0.003) | 0.004 (0.003) | -0.056 (0.005) | -0.012 (0.003) |
| Carnival | 0.016 (0.003) | 0.017 (0.004) | 0.063 (0.004) | -0.015 (0.004) | -0.038 (0.006) | -0.013 (0.003) |
| Easter | 0.013 (0.003) | 0.028 (0.003) | 0.055 (0.003) | -0.052 (0.003) | 0.110 (0.005) | 0.015 (0.002) |
| Father's day | -0.030 (0.003) | 0.011 (0.003) | -0.005 (0.003) | -0.074 (0.003) | 0.086 (0.005) | -0.006 (0.003) |
| Autumn | 0.025 (0.003) | -0.037 (0.003) | -0.047 (0.003) | -0.017 (0.003) | 0.057 (0.005) | -0.004 (0.003) |
| Krokus | -0.005 (0.003) | 0.017 (0.003) | 0.015 (0.003) | 0.034 (0.003) | 0.011 (0.005) | 0.009 (0.002) |
| Liberation day | -0.013 (0.003) | 0.008 (0.004) | 0.023 (0.003) | -0.012 (0.003) | 0.011 (0.005) | 0.003 (0.003) |
| New year | 0.029 (0.003) | -0.054 (0.003) | 0.061 (0.003) | -0.043 (0.004) | 0.009 (0.005) | -0.034 (0.003) |
| Pre-Easter | -0.012 (0.003) | 0.034 (0.003) | 0.113 (0.003) | -0.059 (0.003) | 0.128 (0.005) | 0.019 (0.002) |
| Pre-Pentecost | -0.023 (0.003) | 0.022 (0.004) | 0.013 (0.003) | 0.005 (0.004) | 0.111 (0.006) | 0.033 (0.003) |
| Pre-Christmas | 0.044 (0.005) | -0.010 (0.007) | 0.072 (0.006) | 0.069 (0.006) | 0.085 (0.011) | 0.082 (0.005) |
| Queen's day | -0.006 (0.003) | 0.055 (0.004) | 0.005 (0.003) | -0.019 (0.004) | 0.025 (0.005) | 0.018 (0.003) |
| Mother's day | 0.050 (0.004) | 0.077 (0.005) | 0.007 (0.005) | -0.016 (0.005) | -0.007 (0.007) | -0.026 (0.004) |
| Olympus | 0.096 (0.004) | 0.183 (0.006) | 0.119 (0.004) | 0.062 (0.004) | -0.150 (0.007) | 0.003 (0.005) |
| Pentecost | 0.003 (0.003) | 0.041 (0.003) | 0.064 (0.003) | 0.044 (0.003) | 0.027 (0.005) | 0.027 (0.003) |
| Sinterklaas | 0.054 (0.003) | -0.033 (0.003) | -0.065 (0.003) | -0.026 (0.003) | 0.029 (0.005) | 0.022 (0.003) |
| Christmas | 0.059 (0.004) | -0.003 (0.005) | 0.039 (0.005) | -0.023 (0.005) | 0.138 (0.008) | 0.046 (0.004) |
| Price ratio competition | -0.258 (0.035) | -0.955 (0.023) | -0.292 (0.051) | -0.231 (0.056) | 0.762 (0.059) | -0.162 (0.021) |
| Price ratio regular price | 1.023 (0.043) | -0.481 (0.038) | -3.179 (0.065) | 0.075 (0.052) | 3.353 (0.078) | 0.220 (0.027) |
| Price ratio substitute 1 | -0.731 (0.010) | -0.316 (0.015) | -0.288 (0.033) | 0.052 (0.028) | 1.697 (0.025) | -0.086 (0.023) |
| Price ratio substitute 2 | -0.984 (0.022) | -0.498 (0.013) | 0.239 (0.021) | -0.395 (0.027) | -2.845 (0.025) | -0.188 (0.016) |
| Price ratio complement 1 | -0.017 (0.011) | 0.000 (0.000) | 0.530 (0.024) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| Discount price | -0.005 (0.000) | 0.019 (0.002) | 0.052 (0.001) | 0.016 (0.002) | 0.047 (0.004) | 0.008 (0.002) |
| Trips | 0.004 (0.000) | 0.000 (0.000) | 0.014 (0.000) | 0.018 (0.000) | 0.022 (0.000) | 0.005 (0.000) |
| Instore flyer | -0.362 (0.029) | -1.014 (0.073) | 0.177 (0.039) | -0.347 (0.081) | 0.841 (0.106) | 3.166 (0.269) |
| Online radio | 0.061 (0.020) | 0.000 (0.000) | 0.000 (0.000) | 0.346 (0.022) | -0.517 (0.026) | -0.109 (0.012) |
| Signing on shelf | 0.635 (0.026) | 2.839 (0.064) | -0.339 (0.093) | -0.782 (0.074) | 2.307 (0.205) | -0.154 (0.013) |
| Openings flyer | -1.779 (0.084) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | -1.086 (0.126) | -0.827 (0.067) |
| Store radio | 4.652 (0.326) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| Entrance poster | -2.965 (0.336) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| Second placing | 0.000 (0.000) | 0.552 (0.027) | 1.011 (0.081) | 0.517 (0.068) | 0.347 (0.178) | 0.000 (0.000) |
| National folder | 0.879 (0.082) | -5.877 (0.194) | 0.000 (0.000) | -0.999 (0.119) | 0.000 (0.000) | -1.944 (0.261) |
| Dynamic newsletter | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 1.825 (0.044) | 0.000 (0.000) | 0.000 (0.000) |
| Rain (x1000) | 0.013 (0.003) | -0.066 (0.003) | -0.049 (0.003) | -0.007 (0.003) | 0.079 (0.005) | -0.035 (0.002) |
| Temperature (x1000) | 0.405 (0.008) | 0.630 (0.010) | 0.000 (0.000) | 0.282 (0.011) | -0.040 (0.020) | 0.709 (0.010) |
| Seasonality | 0.152 (0.004) | 0.220 (0.003) | 0.135 (0.004) | 0.088 (0.004) | 0.283 (0.007) | 0.217 (0.002) |
| Percentage high education | -0.212 (0.077) | -0.242 (0.051) | -0.211 (0.052) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| Percentage high income | -0.323 (0.078) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| Franchise dummy | -0.060 (0.020) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | -0.043 (0.015) | -0.036 (0.010) |
| Percentage non-western immigrants | 0.593 (0.221) | 0.839 (0.168) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.369 (0.096) |
| Percentage western immigrants | -1.449 (0.225) | -0.658 (0.212) | 0.000 (0.000) | -0.790 (0.144) | 0.000 (0.000) | 0.000 (0.000) |
| Region South | 0.000 (0.000) | -0.066 (0.016) | -0.166 (0.014) | 0.000 (0.000) | 0.000 (0.000) | -0.043 (0.009) |
| Average family size | 0.000 (0.000) | 0.102 (0.037) | 0.000 (0.000) | -0.144 (0.033) | 0.000 (0.000) | 0.000 (0.000) |
| Percentage unemployed | 0.000 (0.000) | 0.000 (0.000) | -0.858 (0.363) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| City-rural dummy | 0.000 (0.000) | 0.000 (0.000) | -0.036 (0.016) | 0.000 (0.000) | 0.000 (0.000) | 0.044 (0.009) |
| Percentage regular visitor competitor | 0.000 (0.000) | 0.000 (0.000) | -0.093 (0.037) | 0.062 (0.028) | 0.105 (0.033) | 0.000 (0.000) |
| Percentage high social class | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.254 (0.046) | 0.000 (0.000) | 0.000 (0.000) |
| Surface in sqm | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | -0.011 (-0.002) | 0.000 (0.000) | -0.006 (0.001) |
| Self-scan dummy | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.036 (0.014) | 0.043 (0.009) |
| $R^2$ | 0.463 | 0.560 | 0.541 | 0.427 | 0.552 | 0.660 |

**Table C.8:** The table presents the coefficients (and standard errors) of the variables of the panel data regression models on store-item level for all clusters to declare the dependent variable of price elasticities. Note, that the significance level is set to 5%.

| **Panel data regression model**<br>**Independent variables** | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|---|
| Intercept | -0.215 (0.041) | -1.219 (0.038) | -0.532 (0.050) | -0.781 (0.065) |
| Ascension day | 0.052 (0.003) | 0.049 (0.003) | 0.036 (0.004) | 0.043 (0.005) |
| Back to school | 0.006 (0.003) | 0.007 (0.003) | -0.018 (0.004) | -0.009 (0.005) |
| Carnival | -0.003 (0.004) | -0.007 (0.003) | 0.005 (0.005) | 0.001 (0.006) |
| Easter | -0.001 (0.003) | 0.019 (0.003) | 0.008 (0.004) | 0.006 (0.004) |
| Father's day | -0.023 (0.003) | -0.009 (0.003) | -0.038 (0.004) | -0.029 (0.005) |
| Autumn | -0.006 (0.003) | 0.003 (0.003) | -0.006 (0.004) | -0.007 (0.005) |
| Krokus | 0.002 (0.003) | -0.010 (0.003) | 0.018 (0.004) | 0.008 (0.005) |
| Liberation day | -0.015 (0.003) | -0.014 (0.003) | -0.012 (0.004) | -0.020 (0.005) |
| New year | -0.004 (0.003) | 0.004 (0.003) | 0.015 (0.004) | -0.006 (0.005) |
| Pre-Easter | -0.023 (0.003) | 0.044 (0.003) | -0.028 (0.004) | 0.002 (0.005) |
| Pre-Pentecost | -0.009 (0.003) | 0.024 (0.003) | -0.013 (0.004) | 0.005 (0.005) |
| Pre-Christmas | 0.031 (0.006) | 0.073 (0.006) | 0.048 (0.008) | 0.064 (0.009) |
| Queen's day | 0.015 (0.003) | 0.030 (0.003) | 0.013 (0.004) | 0.008 (0.005) |
| Mother's day | 0.016 (0.004) | 0.013 (0.004) | 0.017 (0.006) | 0.009 (0.007) |
| Olympus | 0.030 (0.005) | 0.019 (0.004) | 0.044 (0.006) | 0.028 (0.008) |
| Pentecost | 0.044 (0.003) | 0.036 (0.003) | 0.053 (0.004) | 0.038 (0.005) |
| Sinterklaas | -0.027 (0.003) | -0.001 (0.003) | -0.002 (0.004) | 0.005 (0.005) |
| Christmas | 0.050 (0.004) | 0.065 (0.004) | 0.082 (0.006) | 0.051 (0.007) |
| Price ratio competition | -0.823 (0.054) | -0.197 (0.053) | -1.040 (0.070) | -0.798 (0.095) |
| Price ratio regular price | 1.847 (0.074) | 1.303 (0.067) | 2.197 (0.100) | 1.762 (0.123) |
| Price ratio substitute 1 | -1.083 (0.039) | 0.316 (0.033) | -1.183 (0.049) | -0.387 (0.064) |
| Price ratio substitute 2 | -0.371 (0.038) | -1.165 (0.033) | -0.411 (0.050) | -0.823 (0.060) |
| Price ratio complement 1 | -0.644 (0.027) | -0.477 (0.023) | -0.442 (0.034) | -0.591 (0.044) |
| Discount price | -0.002 (0.002) | -0.011 (0.001) | 0.002 (0.002) | -0.007 (0.003) |
| Trips | 0.013 (0.000) | 0.002 (0.000) | 0.011 (0.000) | 0.006 (0.000) |
| Rain (x1000) | -0.002 (0.001) | -0.015 (0.003) | -0.011 (0.003) | -0.002 (0.001) |
| Temperature (x1000) | 0.000 (0.000) | 0.500 (0.010) | 0.327 (0.015) | 0.430 (0.015) |
| Seasonality | 0.194 (0.005) | 0.205 (0.005) | 0.126 (0.006) | 0.219 (0.006) |
| Instore flyer | -0.446 (0.12) | -0.469 (0.139) | 0.000 (0.000) | -0.786 (0.247) |
| Signing on shelf | 1.227 (0.098) | 0.191 (0.096) | 0.958 (0.059) | 0.686 (0.186) |
| Openings flyer | 1.429 (0.695) | 0.000 (0.000) | 0.000 (0.000) | -0.923 (0.26) |
| Dynamic newsletter | 1.568 (0.138) | 1.463 (0.158) | 0.968 (0.067) | 1.583 (0.266) |
| Entrance poster | -1.998 (0.688) | -13.168 (2.328) | 0.000 (0.000) | 0.000 (0.000) |
| National folder | -1.203 (0.140) | -0.640 (0.120) | 0.000 (0.000) | -0.551 (0.216) |
| Buy-1-get-1-free | -0.618 (0.171) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| Second placing | 0.000 (0.000) | 0.952 (0.130) | 0.000 (0.000) | 0.713 (0.237) |
| Online radio | 0.000 (0.000) | -0.244 (0.029) | 0.000 (0.000) | 0.000 (0.000) |
| Store radio | 0.000 (0.000) | 12.317 (2.326) | 0.000 (0.000) | 0.000 (0.000) |
| Franchise dummy | -0.020 (0.008) | -0.026 (0.010) | 0.000 (0.000) | 0.000 (0.000) |
| No. competitive stores in 15km | -0.003 (0.001) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| Percentage high education | -0.105 (0.035) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| Percentage unemployed | -0.410 (0.202) | 0.789 (0.202) | 0.000 (0.000) | 0.000 (0.000) |
| Self-scan dummy | -0.026 (0.008) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| Percentage regular visitor competitor | -0.050 (0.023) | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) |
| Percentage high social class | 0.000 (0.000) | 0.000 (0.000) | 0.125 (0.042) | 0.000 (0.000) |
| Percentage non-western immigrants | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 1.276 (0.360) |
| City-rural dummy | 0.000 (0.000) | 0.000 (0.000) | 0.000 (0.000) | 0.040 (0.023) |
| $R^2$ | 0.622 | 0.649 | 0.651 | 0.802 |