

Team Assignment 1

Group 6

February 28, 2017

Contents

1. Introduction	1
2. Methodology	1
2.1. Data Transformation	1
2.2. Regression Model	2
3. Results and Discussion	2
4. Conclusion	2
Appendices	2

1. Introduction

This report aims to understand the market of four chosen fast-moving consumer goods (FMCG) products in the Italian market. Two product categories, i.e. Crackers and Dry Cookies, are provided in the dataset. Of the 17 brands, four brands from the Crackers category are chosen to be analysed further. In general, to obtain further information about the market of those products, detailed information of the demand elasticity of each product has to be extracted from the dataset. The method to capture such information will be explained in the next section. Then, the results will be discussed extensively in the following part to understand the structure and competitiveness of the market of the four brands (the market for crackers?).

2. Methodology

Our methodology will be divided into 2 parts, data transformation and regression model. Due to limited information on the dataset, we will need to perform several transformation to the data. The following part will explain the model used in the report. Linear regression models are predominantly chosen to make an inference about the relationship between price and quantity.

2.1. Data Transformation

The objective of this part is to provide the models with the most appropriate features and to separate the effect of different demand factors of each product. Prices of each brand are derived from dividing the sales value of each week by the sales volume. Since the relationship between

volume and price is not known, the two variables should be converted to their log forms to adjust for models with different functional forms.

The next step is to separate the effect of promotion and observe when promotion occurred for each brand. However, **Product 2** and **Product 5** data does not have Gross Rating Point (GRP) variable which would tell the size offline advertising corresponding to the product. Another parameter is needed to see when a promotion campaign take place. The first method is to determine the week when the products are selling at discounted price. We compute the average prices (\bar{P}_i) and their standard deviation (σ_i) over 105 weeks. If the price goes below $\bar{P}_i - \sigma_i$, the **promo** variable is coded 1 during that week and 0 if otherwise. Then, the effect of GRP could also be transformed into binary variable to simplify the impact of promotion and discount on the sales volumes.

After separating all possible factors that could affect the price and volume, we need to verify the effect of promotion and discount for each product. This can be obtained by calculating the baseline sales volume and the lift. We estimate the weeks where there were advertising campaign for all products, particularly the two products without GRP data, to get the best representation of when the lift occurred.

{should explain the way we calculate baseline price & lift here?}

2.2. Regression Model

Firstly, To get the elasticity of each product and its cross elasticity against other products, we computed three models with different functional forms (Linear, Semi-Log, and Log-Log). **Volume** of the evaluated brand is the dependent variable and its **Price** and promotions (**ad** and **promo**) are our first three independent variables. Then, the prices of other products and their promotion campaigns are added to the model to capture the cross-elasticity of the selected product. The process is identical for different functional forms. In general, the Full Model for **Product 2** looks like the following, but the Reduced Model are different for each product depending on the significant variables.

$$\widehat{volume_2} = price_2 + ad_2 + promo_2 + \sum_{i=3}^5 (price_i + ad_i + promo_i)$$

Secondly, stepwise selections are needed to omit non-significant variables from the model to get better estimation. Then, the three different functional models of each brand are compared against each other by computing their root-mean-square error (RMSE) or the difference between predicted values of **Volume** and the observed values. The model with the lowest RMSE is selected to calculate the elasticity and cross-elasticity of the product. Subsequently, the elasticity matrix can be obtained.

3. Results and Discussion

4. Conclusion

Appendices