Team Assignment 1

Group 6

Contents

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, i.e. Product 2, Product 3, Product 4, and Product 6, are selected to be analyzed further. The first step will be to conduct a preliminary analysis of each product. This will allow to form initial hypotheses about the different positions of the brands in the market as well as the market structure. After the initial analysis, we will attempt to detect the promotional periods using a combination of algorithmic and visual apporach. Then, the own and cross-price elasticities will be calculated to assess the clout and vulnerability of each brand. Finally, the report will conclude with a discussion of the results and insights gained abouut the Italian cracker market.

2. Preliminary Analysis

Before engaging in elasiticity calculations, we aim to gain information about the products as they are presented in the data. Table 1 shows the average values for volume and price for each cracker brand. Initial inspection reveals that Product 2 is the price leader with the lowest average price and sales volume. Product 3 has slightly lower weekly sales despite its higher price. The product could be from a mid-range segment or have different packaging options. Product 3 and Product 6 have similar average prices although significantly different volumes. Product 4 appears to be the premium product amongst the four brands with relatively higher average price and volume.

	Average Price	Average Volume
Product 2	2.663	94527.33
Product 3	5.278	84788.64
Product 4	6.737	47015.71
Product 6	5.311	33132.16

Table 1: Average price and sales volume for each brand.

With this preliminary analysis we expect to see brands with lower prices have a relatively higher own price elasticity than brands with higher prices. We also expect brands with relatively lower prices to be more vulnerable and mid-ranged brands to have the highest clout.

3. Promotion Detection

In order to calculate elasticites, we need to isolate the effect of promotions on volume. The first apporach used to detect promtional periods was to use the average price ($\bar{P}_i$) and standard deviation (σ_i) of each brand to create a price promotion threshold. Any period with a price below one standard deviation from the average was seen as a promotional period. This method of promotion detection was supplemented with a visual approach. Here, periods that showed high sales volumes were also considered promotional periods. Rather than being price promotions, these are likely to be special end of aisle displays or features the brands engage in. If included in the data set, gross rating point (GRP) was encoded advertising period rather than promotional period.

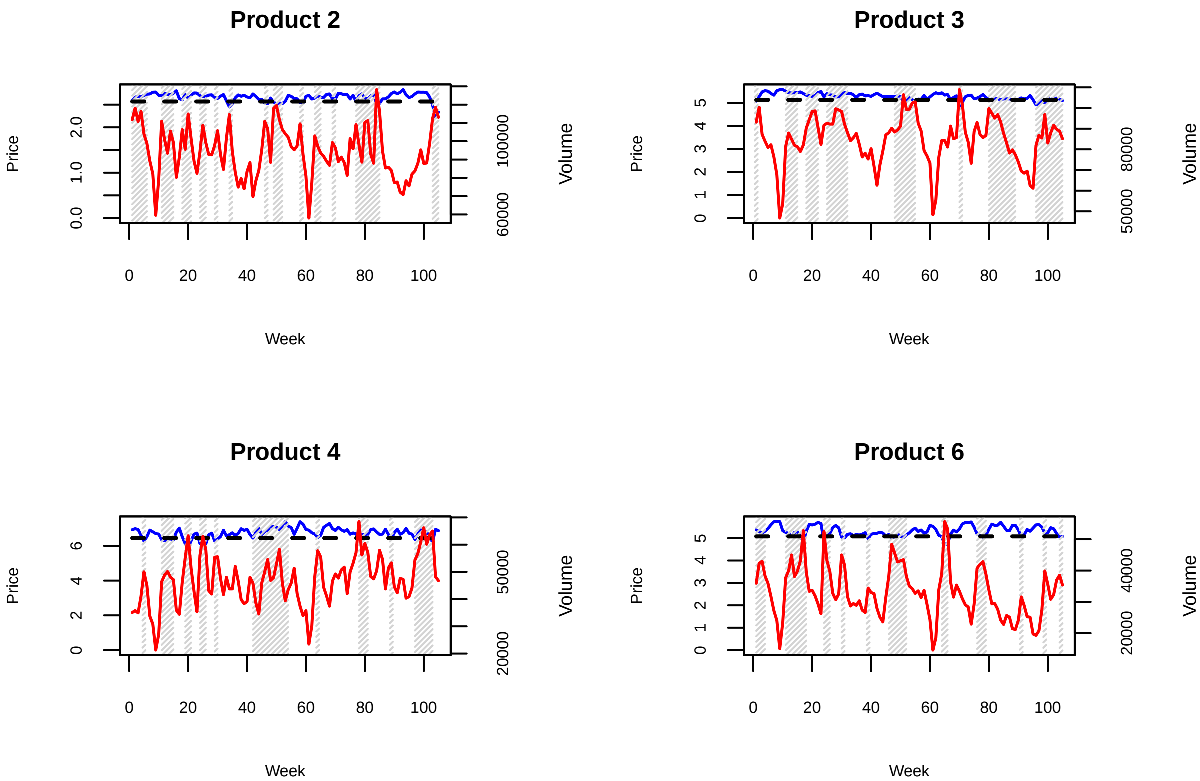

Chart 1: Price (blue), sales volume (red), promotion (gray), price promotion threshold (black dashed). Note: The y-axes are of different scales.

The four different brands display different promotional behaviors (above). Product 2, for example, engages in many short promotional periods which were identified as flash sales. Product 3 appears to favor longer periods of promotion. Product 4 and Product 6 display a mixture of both behaviors. The effectiveness of the respective brands promotions vary greatly across the given time period with some promotions boosting sales by as much as 40% and others by as little as 15%.

4. Baseline and Lift Calculation

After the promotional periods are identified we calculate the volume baseline and the lift for each brand. This is important to isolate the effect of promotions on sales. In order to do so, we remove promotional periods from the data. Although information about the competition was available, it was decided that the models to use for the baseline calculations should only contain information about the brand itself. This was done to calculate the pure effect of a brand's own promotion. The baseline volume was then calculated using a linear regression in combination with a stepwise algorithm to determine the best combination of features. Using the calculated baseline we derived the lift by subtracting the baseline from the volume for each period.

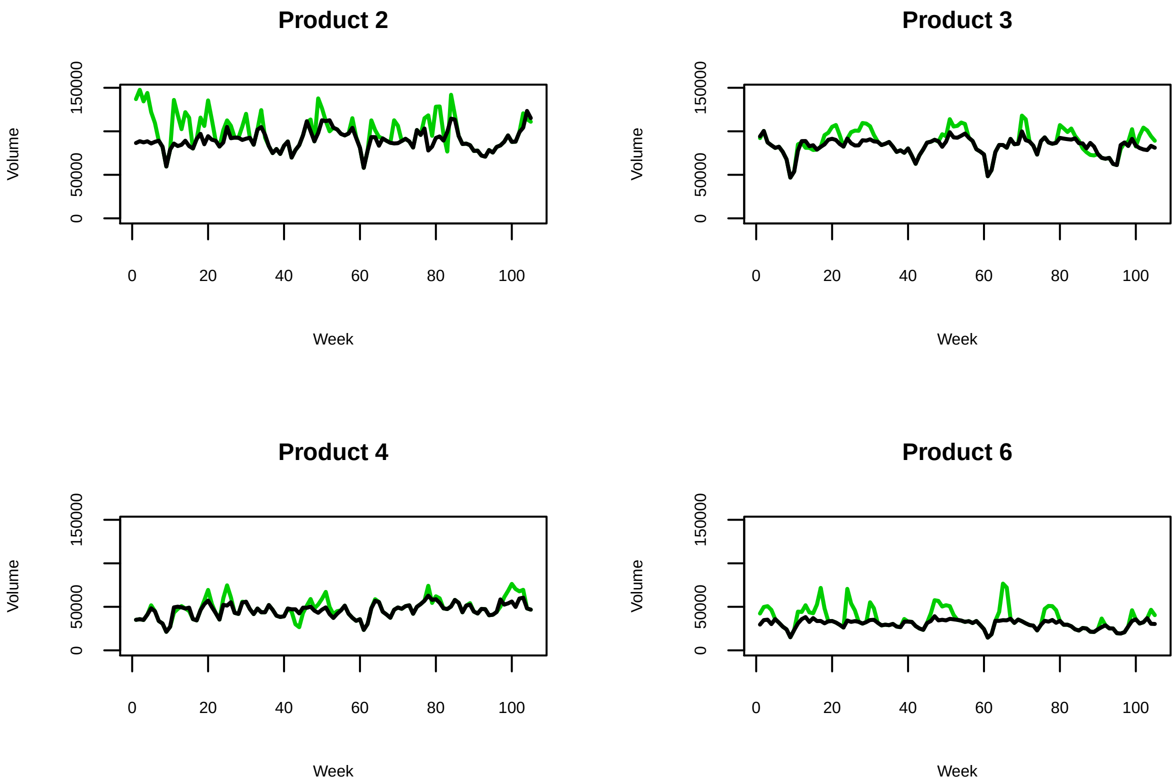

Chart 2: Sales volume (black), lift (green), promotion (gray). Note: The y-axes are of the same scale.

The four brands show different effectiveness in their promotions (above). Product 2 has the most successful promotional periods with large added volume to its baseline sales. This stands in contrast to Product 3 and Product 4, which see relatively smaller efffectiveness from their promotions. Product 6 sees the most balanced effect from promotions. This is likely due its the relatively regular baseline sales compared to other brands.

5. Regression Model

To calculate own and cross price elasticities for each brand we built models with different functional forms (linear, semi-log, and log-log). The dependent variable is `volume` and `price`, `ad` (advertizing) and `promo` (promotion) are the first three independent variables. Then, information about the competition was added to account for their potential effects. For the Semi-Log form we use the log of `volume` as the independent variable while the predictors stay the same. For the log-log model, we use the log of `volume` as the independent variable as well as log forms for `price`. The general models look as follows:

Linear:

$$volume_i = \quad a_0 + a_1 price_i + a_2 ad_i + a_3 promo_i + \sum_j (b_1 price_j + b_2 ad_j + b_3 promo_j)$$

Semi-log:

$$log(volume_i) = \quad a_0 + a_1 price_i + a_2 ad_i + a_3 promo_i + \sum_j (b_1 price_j + b_2 ad_j + b_3 promo_j)$$

Log-log:

$$log(volume_i) = \quad a_0 + a_1 log(price_i) + a_2 ad_i + a_3 promo_i + \sum_j (b_1 log(price_j) + b_2 ad_j + b_3 promo_j)$$

For each model, every combination of predictors was evaluated using a stepwise selection algorithm. The best models for each functional form were then compared by their root mean squared error (RMSE). The model with the lowest RMSE was selected to calculate own and cross-price elasticities.

6. Results and Discussion

The best model for Product 2 and Product 6 were the linear models. The best models for Product 3 and Product 4 were semi-log and log-log respectively. The formulas used to calculate own and cross-price elasticities given the selected model are as follows:

Linear:

$$e_{ii} = a_1 \times \frac{\bar{P}_i}{\bar{Q}_i}$$
$$e_{ij} = b_1 \times \frac{\bar{P}_j}{\bar{Q}_i}$$

Semi-log:

$$e_{ii} = a_1 \times \bar{P}_i$$
$$e_{ij} = b_1 \times \bar{P}_j$$

Log-log:

$$e_{ii} = a_1$$
$$e_{ij} = b_1$$

Table 2: Elasticity Matrix

	Product 2	Product 3	Product 4	Product 6
Product 2	-1.99	0.000	-0.328	1.026
Product 3	0.00	-2.094	0.000	0.275
Product 4	0.00	0.000	-3.020	0.000
Product 6	0.00	0.000	0.000	-2.319

When examining the elasticity matrix, we notice a number of interesing things. The own price elasticites display a range from -3.020 to -1.99. These results are in line with what is expected in the food FMCGs. Product 4 has the lowest price elasticity at -3.02. As the premium brand, its sales are more price sensitive as consumers already pay a high price. Product 2 has the highest own price elasticity at -1.99. This implies that, within the scope of our analysis, buyers of Product 2 are the least price sensitive.

With regards to the cross price elasticities, many of the products have no effect on each other. Notable cross price elasticities include Product 6's effect on Product 3. As two similarly priced products, when Product 6 increases its price, Product 3's sales increase. This makes sense as customers often equate similarly priced products to direct substitutes. A similar, although more pronounced effect is seen between Product 6 and Product 2. When Product 6's price increases, the increase in sales in Product 2 is bigger than the one one in Product 3. Reasons for the increased effect could be shelf placement or simplified customer decision rule to revert to the cheapest brand when the price of their favorite brand goes up. Another noticeable effect is the unexpected negative effect of Product 4 on Product 2. This means that Product 4, the premium cracker, and Product 2, the price leader, can be seen as complementary goods. As Product 4's price increases, Product 2's sales volume decreases. This negative cross price elasticity could also arise due to inconsistencies when aggregating the data to a weekly level. The aggregation would mask sales and promotions that could only be seen on a more detailed level, such as with daily store level sales data.

When analyzing the clout and vulnerability plot (Chart 3), we see that Product 2 has low clout and high vulnerability. This confirms the intuition that the price leader is vulnerable to changes in prices from higher priced competitors and that it has little effect on competitors sales. This product competes solely on price. Product 3 has a higher clout and lower vulnerability. As the standard offering, it can somewhat affect the sales of other products but is also vulnerable to their price changes. Product 4, the premium offering has a very low vulnerability and also low clout. Because no crackers operate in the premium market it is not vulnerable to poaching customers. Premium customers will usually buy Product 4 no matter what. Conversely, because it is a premium product it does not affect the sales of other products. Finally, Product 6 has a high clout and high vulnerability. As a niche offering, its price changes can alter its sales volumes significantly. These changes will directly affect the lower priced offerings in the market. However, it is also vulnerable because people will only purchase a niche cracker when it has a notable price advantage relative to the standard offerings.

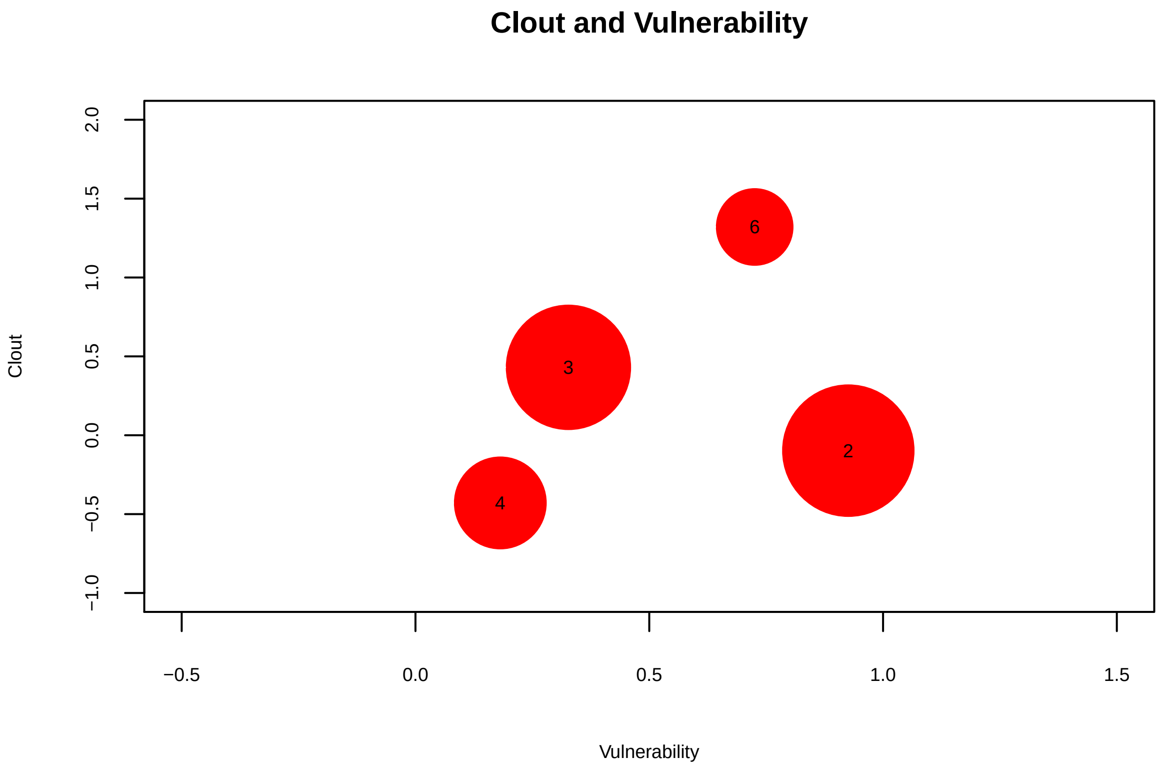

Chart 3: Bubble size represents the average sales volume. Labels are the respective brands.

7. Conclusion

The analysis of the Italian cracker market, through the selection of 4 brands, revealed several interesting results. The promotional strategies vary significantly across brands. While Product 2 engages in several short promotions, Product 3 favors longer periods. Products 4 and 6 use a mix of short and long promotions. The effect of the respective Products' promotions also varies greatly. Whereas Product's 2, 3, and 4 have similar variation in their baseline sales, Product 2 is a lot more efficient in promoting made obvious by the significant lift. Product 6 has the most stable baseline sales out of all four.

The analysis of clout and vulnerability for Product 2 shows the typical behavior of a price leader. The product has no effect on its competitors' sales but is in turn very vulnerable to what it's competitors do. Product 4 also displays the behavior expected from a premium product. It is not very vulnerable and has low clout as it operates in a segment of its own. Product 6 exhibits a mixed behavior.

The high promotional intensity shows that the market for crackers is competitive and offers little options for differentiation. The market shows no signs of apparent seasonality and no signs of growth. As such, based on this analysis it is not an attractive for new players.

Appendices

Table 3: Product 2 Regression Models

	Dependent variable:		
	Volume_2	log.q_2	
	Linear	Semi-Log	Log-Log
	(1)	(2)	(3)
promo_2	6,510.87***	0.05***	0.05***
	(1,804.06)	(0.02)	(0.02)
price_2	−70,639.57***	−0.79***	
	(8,400.99)	(0.10)	
Average_Distribution_6	1,050.68**	0.01***	0.01***
	(456.68)	(0.005)	(0.005)
log.p_2			−2.02***
			(0.25)
Average_Distribution_3	2,599.02***	0.03***	0.03***
	(596.64)	(0.01)	(0.01)
price_6	18,263.11***	0.18***	
	(3,576.79)	(0.04)	
Average_Number_SKUs_2	59,635.95***	0.59***	0.61***
	(15,697.33)	(0.17)	(0.17)
log.p_6			0.92***
			(0.21)
promo_6	6,103.97***	0.05**	0.05**
	(1,860.52)	(0.02)	(0.02)
log.p_4			−0.42**
			(0.18)
Average_Number_SKUs_6	56,587.17***	0.77***	0.79***
	(17,270.68)	(0.19)	(0.20)
price_4	−4,605.24*	−0.06**	
	(2,438.71)	(0.03)	
ad_3		−0.03*	−0.03*
		(0.02)	(0.02)
Constant	−353,958.70***	6.02***	5.67***
	(72,148.96)	(0.79)	(0.89)
Observations	105	105	105
R^2	0.83	0.84	0.83
Adjusted R^2	0.81	0.82	0.82
Residual Std. Error	5,906.55 (df = 95)	0.06 (df = 94)	0.06 (df = 94)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 4: Product 3 Regression Models

	Dependent variable:		
	Volume_3	log.q_3	
	Linear	Semi-Log	Log-Log
	(1)	(2)	(3)
Average_Distribution_3	2,081.42**	0.03***	0.03***
	(793.07)	(0.01)	(0.01)
Average_Distribution_6	1,668.52***	0.02***	0.02***
	(482.94)	(0.01)	(0.01)
price_3	−34,030.91***	−0.40***	
	(4,358.26)	(0.05)	
log.p_3			−2.07***
			(0.27)
Average_Number_SKUs_6	38,545.69***	0.41***	0.41***
	(12,872.54)	(0.15)	(0.15)
promo_3	3,205.35***	0.03**	0.03**
	(1,050.48)	(0.01)	(0.01)
Average_Number_SKUs_2	27,912.28**	0.34**	0.34**
	(11,488.41)	(0.14)	(0.14)
Average_Distribution_4		0.02*	0.02*
		(0.01)	(0.01)
price_6	4,438.68**	0.05**	
	(2,060.98)	(0.02)	
Average_Number_SKUs_3	28,558.72***	0.37***	0.38***
	(7,438.81)	(0.09)	(0.09)
week_Number_2	−120.78***	−0.002***	−0.002***
	(40.06)	(0.001)	(0.001)
ad_4	2,264.40**		
	(1,008.80)		
promo_2	1,983.24**	0.02*	0.02*
	(962.19)	(0.01)	(0.01)
log.p_6			0.26**
			(0.13)
Constant	−335,415.40***	3.81***	4.99***
	(71,350.30)	(0.90)	(1.03)
Observations	105	105	105
R^2	0.88	0.90	0.90
Adjusted R^2	0.87	0.89	0.89
Residual Std. Error (df = 93)	4,181.33	0.05	0.05
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 5: Product 4 Regression Models

	Dependent variable:		
	Volume_4	log.q_4	
	Linear	Semi-Log	Log-Log
	(1)	(2)	(3)
Average_Distribution_4	1,079.59*	0.04***	0.04***
	(552.58)	(0.01)	(0.01)
price_4	−20,400.14***	−0.45***	
	(1,386.12)	(0.02)	
promo_4	1,059.29		
	(766.40)		
log.p_4			−3.02***
			(0.16)
Average_Number_SKUs_3			0.14
			(0.11)
Average_Number_SKUs_4	22,548.64***	0.62***	0.59***
	(5,590.48)	(0.08)	(0.08)
Average_Distribution_6	930.19**	0.02***	0.02***
	(386.93)	(0.01)	(0.01)
ad_3	−1,757.62**	−0.03*	−0.04**
	(757.30)	(0.02)	(0.01)
Average_Number_SKUs_6	35,722.50***	0.61***	0.55***
	(8,621.82)	(0.15)	(0.16)
ad_4	1,532.88**		
	(710.51)		
week_Number_2	74.57*		
	(42.51)		
Average_Number_SKUs_2	19,109.04**	0.48***	0.40**
	(8,163.03)	(0.16)	(0.16)
Average_Distribution_3		0.02*	0.02*
		(0.01)	(0.01)
Average_Distribution_2	727.21**	0.01	0.01**
	(330.42)	(0.01)	(0.01)
ad_6	−1,010.91	−0.03**	−0.03*
	(671.99)	(0.01)	(0.01)
promo_3		0.02	
		(0.02)	
Constant	−236,882.30***	1.76*	4.51***
	(43,226.01)	(0.93)	(0.99)
Observations	105	105	105
R^2	0.91	0.94	0.94
Adjusted R^2	0.90	0.93	0.93
Residual Std. Error	2,885.84 (df = 92)	0.06 (df = 93)	0.06 (df = 93)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 6: Product 6 Regression Models

	Dependent variable:		
	Volume_6	log.q_6	
	Linear	Semi-Log	Log-Log
	(1)	(2)	(3)
promo_6	5,175.60***	0.13***	0.13***
	(921.89)	(0.03)	(0.03)
Average_Number_SKUs_6	43,603.82***	1.39***	1.39***
	(8,135.19)	(0.24)	(0.24)
price_6	−14,469.99***	−0.40***	
	(1,837.43)	(0.05)	
log.p_6			−2.12***
			(0.27)
ad_3	−3,262.87***	−0.08***	−0.08***
	(794.19)	(0.02)	(0.02)
Average_Distribution_6	650.81**	0.04***	0.04***
	(278.58)	(0.01)	(0.01)
Average_Distribution_4	1,337.09***	0.05***	0.05***
	(449.01)	(0.01)	(0.01)
ad_6	−946.65	−0.04**	−0.04**
	(649.80)	(0.02)	(0.02)
ad_4	1,121.77		
	(724.30)		
Average_Number_SKUs_4	−4,322.23		
	(3,139.00)		
Constant	−152,691.50***	1.63*	3.07***
	(33,480.00)	(0.91)	(1.00)
Observations	105	105	105
R^2	0.88	0.89	0.89
Adjusted R^2	0.87	0.88	0.88
Residual Std. Error	2,986.59 (df = 95)	0.09 (df = 97)	0.09 (df = 97)
Note:			*p<0.1; **p<0.05; ***p<0.01