Team Assignment 1

Group 6

February 28, 2017

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 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. Preliminary Analysis

The table below shows the average figures for each Crackers brand. By looking at those information, we assume that `Product 2` is the price leader as it has the lowest price and highest average sales volume. `Product 3` has slightly lower weekly sales despite it higher price. The product might be from more middle segment or has a bigger package and net weight. `Product 2` and `Product 6` could be in the same segment and they are direct competitors because they have similar prices. `Product 4` could be the more premium product amongst the four brands. Those are the initial hypothesis that could be extracted by above information. The analysis that follows this section will capture more comprehensive information from the data.

Table 1: 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

3. Promotion Detection

The objective of this part is to separate the effect of promotion and observe when promotion occured for each brand. However, `Product 2` 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.

4. Baseline and Lift Calculation

{Sally's part}

5. 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 campagins 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.

$$\widehat{volume}_2 = \quad a_0 + a_1 price_2 + a_2 ad_2 + a_3 promo_2 + \textstyle\sum_{i=3}^{6}(b_1 price_i + b_2 ad_i + b_3 promo_i)$$

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.

6. Results and Discussion

After running all model with different functional forms, different functional form models have the lowest RMSE for different products. `Product 2` and `Product 6` linear models are the most efficient when the log form are converted back to level form. We should calculate the average price and quantity of each product over 105 weeks and compute the elasticities. The formula used to obtain these would be the following.

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

The Semi-Log model for `Product 3` has the lowest RMSE amongst the three. The price elasticity formula is expressed below.

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

The best model for `Product 4` is the Log-Log model. Hence, the elasticities are the coefficient of the corresponding price variables. Then, the elasticity matrix and the clout and vulnerability of each product are shown below.

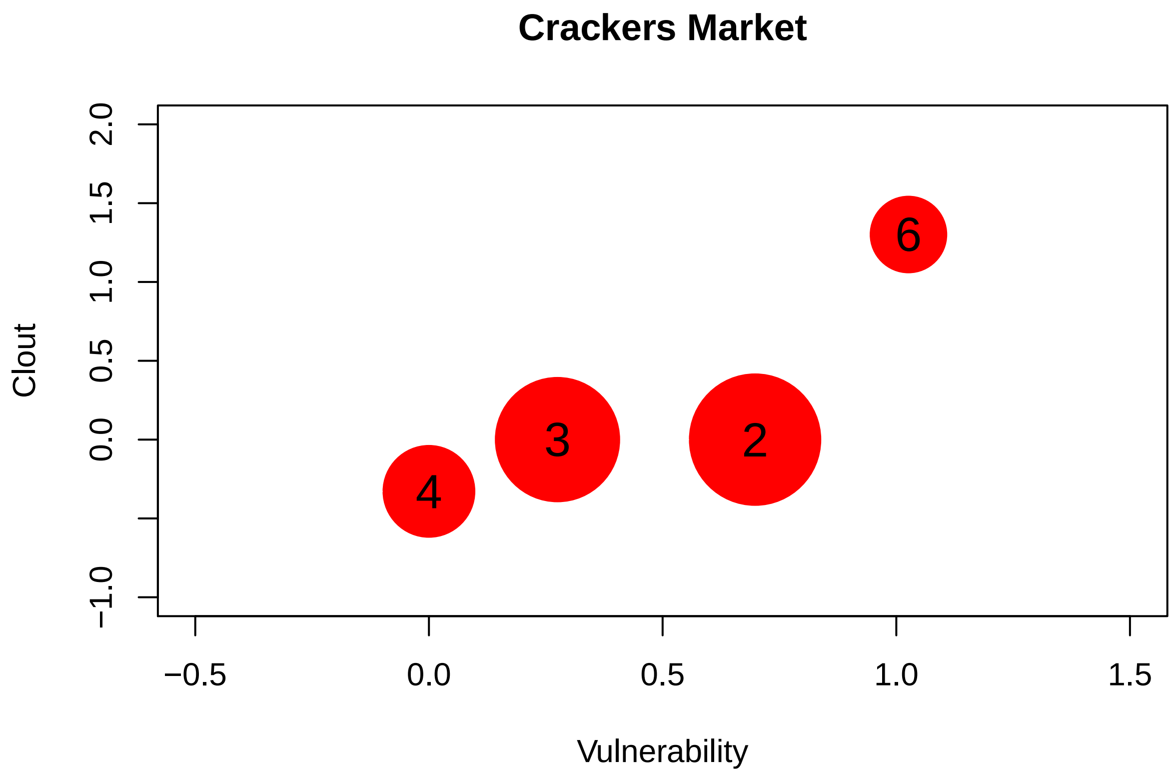
Table 2: Elasticity matrix of Crackers brands

	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

Table 3: Clout and Vulnerability

Product	Vulnerability	Clout
2	0.6978265	0.0000000
3	0.2750607	0.0000000
4	0.0000000	-0.3282258
6	1.0260523	1.3011130

The price elasticity of each product seems suitable for FMCG products. Their demand are elastic as expected from the market where there are many alternatives. However, the market has different products that appeals several customer segments. Usign the clout and vulnerability table, we could locate the position of each brand within the market.



As expected, `Product 2` has the highest vulnerability because it seems to be the cheapest brand in the market and it has no power to affect other brands if it lower the price. However, if the other products lower their prices, customers are more likely to switch to different brands with better quality. `Product 3` seems to be the standard brand-name cracker that everyone loves. Price change does not affect other products although it is quite vulnerable to `Product 6` change of price as they seem to appeal to the same segment. However, `Product 6` price changes have the biggest impact on `Product 2`. Although their average price difference is wide, `Product 6` seems to be the most probable

7. Conclusion

Appendices

Table 4: 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 5: 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 6: 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 7: 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