

# Team Assignment 1

*Group 6*

## 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 its 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

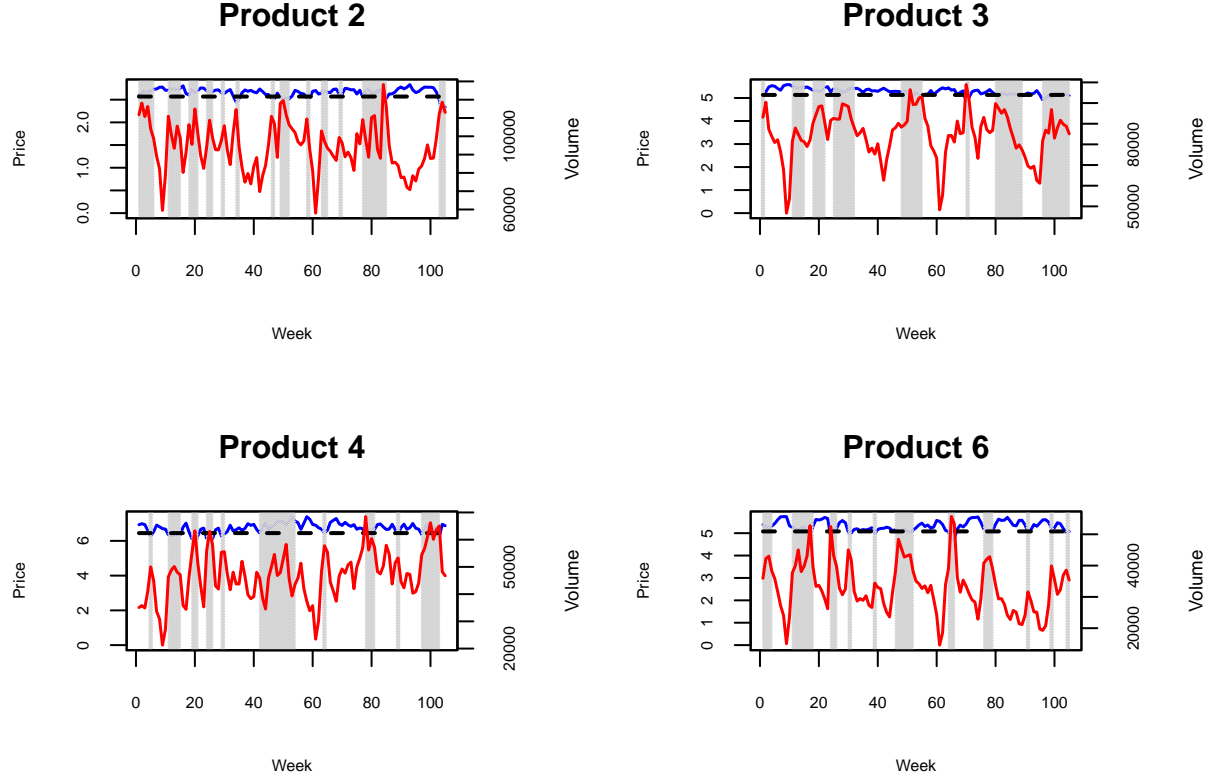
	Average Price	Average Volume
Product 2	2.663	94527.33
Product 3	5.278	84788.64
Product 4	6.737	47015.71
Product 6	5.311	33132.16

## 3. Promotion Detection

The objective of this part is to separate the effect of promotion and observe when promotion occurred for each brand. However, **Product 2** does not have Gross Rating Point (GRP) variable which would tell the size of offline advertising corresponding to the product. Another parameter is needed to see when a promotion campaign takes 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 ( $\sigma_i$ ) 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.

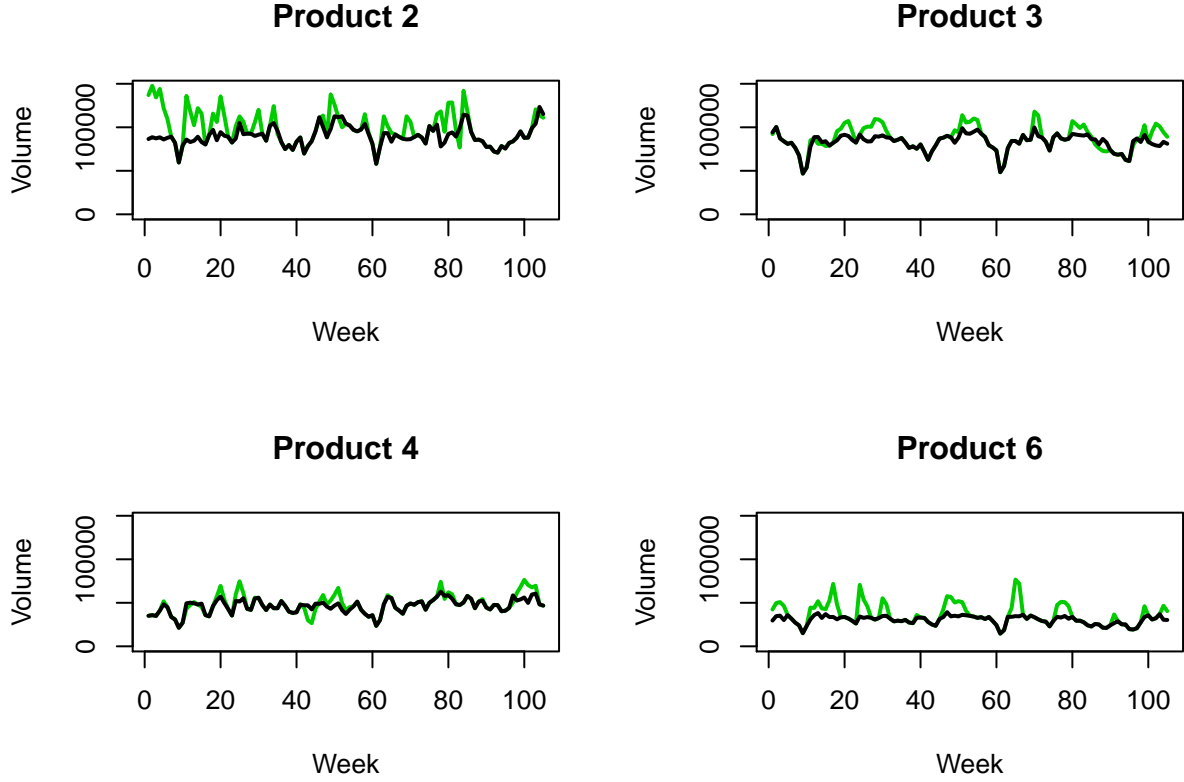


## 4. Baseline and Lift Calculation

Volume baseline and the lift in sales are calculated for each of the product. For each product, filter the data to have only non promotional periods containing volume of sales and all the relevant factors affecting it (week number, average number of SKUs, average distribution, and price of product).

The four relevant factors affecting volume of sales are being regressed to the volume of sales by a linear model. However, only significant factors need to be included in the model hence automatic model selection are performed to select best linear model. Lastly, using the best model selected, volume baseline is derived by putting the values of independent variables into the model.

Besides volume baseline, for promotional periods there are usually lift in sales. Lift in the volume of sales is basically indicates how much the volume of sales increased from its baseline as an effect of promotions. Hence lift is calculated by subtracting actual volume with volume baseline.



## 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 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.

$$\widehat{volume_2} = a_0 + a_1 price_2 + a_2 ad_2 + a_3 promo_2 + \sum_{i=3}^6 (b_1 price_i + b_2 ad_i + b_3 promo_i)$$

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.

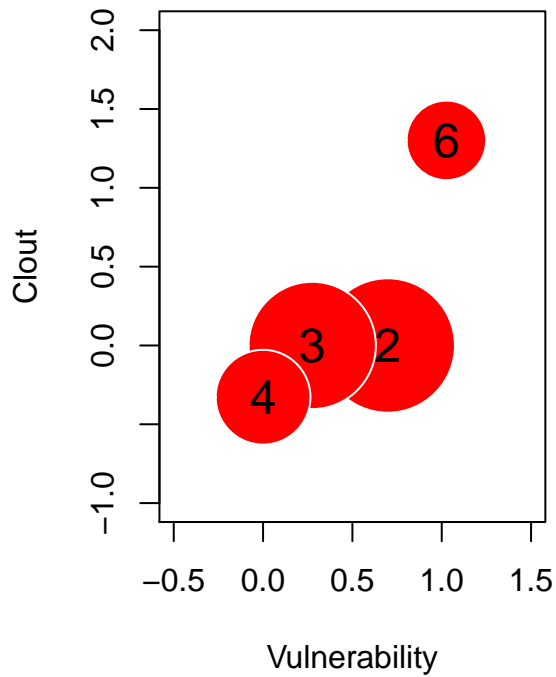
$$e_{ii} = a_1 \times \bar{P}_i$$
$$e_{ij} = b_1 \times \bar{P}_j$$

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.

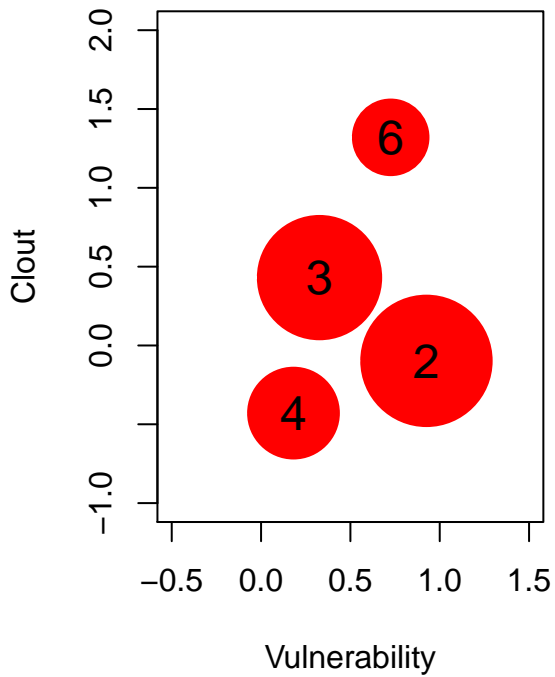
The price elasticity of each product seems suitable for FMCG products. Their demand are elastic as expected from the market where there are many alternatives. However, the market has different products that appeals several customer segments. Usign the clout and vulnerability table, we could locate the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

	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
	Product 2	Product 3	Product 4	Product 6
Product 2	-2.223	0.353	-0.323	0.896
Product 3	-0.016	-2.169	-0.018	0.362
Product 4	0.002	0.117	-2.909	0.063
Product 6	-0.083	-0.040	-0.088	-2.320

**Crackers Market**



**Crackers Market with all Variable**



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

When examining the elasticity matrices, we notice a number of trends. In the elasticity matrix calculated with all potential variables, even those that are insignificant at a 5% significance level, we notice that products 2, 3, and 6 have similar elasticities. With a price elasticity of around -2.2, these

crackers are quite elastic to demand. Product 4 has a much higher price elasticity at -2.9087613. As a more premium good, its sales are more price sensitive as consumers may be inclined to switch to it when it does not have as much of a price premium. Conversely, when product 4 increases its price, customers do not buy it because they cannot justify the even higher price premium.

With regards to the cross elasticities, many of the products have not effect on each other as their prices change. While there exist a few negative cross elasticities, most of them are essentially 0 so we can treat them as such. Notable cross elasticities include product 6's effect on product 3. As two similarly priced products, when the niche product 6 lowers in price, product 3's sales suffer. Customers may be incentivized to try the niche product at a lower price point, compared to the similarly priced standard crackers. A similar effect occurs with product 3's effect on product 2. When product 3's price drops, the low price option of product 2 has less sales. Customers are willing to upgrade to a higher priced product when the price difference is not as much. Product 6 has a very strong effect on the volume of product 2 as well. This is the same effect as with product 3, but not the additional niche factor of product 6 comes into play. Customers are quite likely to try the niche crackers when they are not much more expensive than the standard cheap crackers. Finally, it is interesting to note that product 4's effect on product 2 is actually negative, with a cross elasticity of -0.3234974. This means that as product 4, the expensive crackers, and product 2, the cheap crackers, can be seen as somewhat complementary goods. As product 4's price decreases, product 2's sales volume actually increases. This could potentially make sense because people see the expensive cracker on sale, but upon further examination end up purchasing the cheap cracker. The 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.

In the elasticity matrix calculated when we discard insignificant predictors, the general picture remains the same. Product 2, the lowest priced product in our analysis, has the lowest elasticity at -1.9902055. Product 4, the highest priced, has the highest elasticity at -3.0199726. The only non-zero cross elasticities are product 6's effect on product 3, where people switch to the niche product when it becomes cheaper than the standard offering. The other non-zero cross elasticity is product 4's effect on product 2. Since it is negative, when product 4 decreases in price people are more inclined to buy crackers as a whole. When such customers actually decide which cracker to buy, they end up choosing the cheapest option. Therefore, sales of product 2 are boosted.

When analyzing the clout and vulnerability matrices, we examine both the scenario where we include insignificant predictors and the one where only significant predictors are included. When we include all predictors, we see that product 2 has low clout and high vulnerability. This makes sense because it is a low priced good that people only buy when they cannot afford the more expensive options. 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 high priced offering has a very low vulnerability and also low clout. Because no crackers operate in the premium market is is not vulnerable to poaching customers. Premium customers will only buy product 4. Conversely, because it is a premium product it cannot change the sales of the other products because only premium customers buy product 4. 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 quite vulnerable because people will only purchase a niche cracker when it has a notable price advantage relative to the standard offerings.

The clout / vulnerability matrix looks quite similar when we only include significant predictors in

the analysis. All products have low clout in the market except for the niche product 6. This is the only product that affects others' sales volumes. People will switch to the niche offering when it has a low price premium compared to standard products. Additionally, the lower priced products are more vulnerable to others' price changes. Since they compete on price rather than differentiation, when the competition changes its pricing structure, low priced goods suffer.

## **7. Conclusion**

# Appendices

Table 2: Product 2 Regression Models

	<i>Dependent variable:</i>		
	Volume_2	log.q_2	
	Linear	Semi-Log	Log-Log
	(1)	(2)	(3)
promo_2	6,510.87*** (1,804.06)	0.05*** (0.02)	0.05*** (0.02)
price_2	-70,639.57*** (8,400.99)	-0.79*** (0.10)	
Average_Distribution_6	1,050.68** (456.68)	0.01*** (0.005)	0.01*** (0.005)
log.p_2			-2.02*** (0.25)
Average_Distribution_3	2,599.02*** (596.64)	0.03*** (0.01)	0.03*** (0.01)
price_6	18,263.11*** (3,576.79)	0.18*** (0.04)	
Average_Number_SKUs_2	59,635.95*** (15,697.33)	0.59*** (0.17)	0.61*** (0.17)
log.p_6			0.92*** (0.21)
promo_6	6,103.97*** (1,860.52)	0.05** (0.02)	0.05** (0.02)
log.p_4			-0.42** (0.18)
Average_Number_SKUs_6	56,587.17*** (17,270.68)	0.77*** (0.19)	0.79*** (0.20)
price_4	-4,605.24* (2,438.71)	-0.06** (0.03)	
ad_3		-0.03* (0.02)	-0.03* (0.02)
Constant	-353,958.70*** (72,148.96)	6.02*** (0.79)	5.67*** (0.89)
Observations	105	105	105
R <sup>2</sup>	0.83	0.84	0.83
Adjusted R <sup>2</sup>	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 3: Product 3 Regression Models

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

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 4: Product 4 Regression Models

	<i>Dependent variable:</i>		
	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 <sup>2</sup>	0.91	0.94	0.94
Adjusted R <sup>2</sup>	0.90	0.93	0.93
Residual Std. Error	2,885.84 (df = 92)	0.06 (df = 93)	0.06 (df = 93)

Note:

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 5: Product 6 Regression Models

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

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01