Case Study No. 2 - GoodBelly: Investigating the Relationship Between Store Promotions and Sales

Alice Zhao, Valentin Vrzheshch

October 14, 2016

Executive Summary

The purpose of this case study is to investigate the relationship between promotional activities of demo and endcap and GoodBelly sales. This case study quantitatively addresses the question by building multiple linear regression model to analyze data from 126 Whole Foods stores over the 10 weeks between May 4 and July 13. Based on total of 1,386 observations we developed the model that justifies the demo and endcap activities and shows that those expenditures increase sales. The model is chosen by minimization of BIC criterion, and it has no multicollinearity issue and its residuals are homoscedastic, normally distributed and serially uncorrelated.

1 Introduction

1.1 Data Summary

The data set provides nine time series variables (see tab.1 below).

	Variable	Description
1	**Weekly Sales (Volume)**	The number of units sold per store per week.
2	**Average Retail Price**	The average retail price for GoodBelly products per store per week.
3	**Sales Rep**	Defined as 1 if the store had a regional sales rep (face-to-face contact).
4	**Endcap**	Defined as 1 if a store participated in an endcap promotion.
5	**Demo**	Defined as 1 if the store had a demo on the corresponding week.
6	**Demo1-3**	Defined as 1 if the store had a demo 1-3 weeks ago.
7	**Demo4-5**	Defined as 1 if the store had a demo at least 4-5 weeks ago.
8	**Natural Retailers**	The number of other natural retailers within 5 miles of each store.
9	**Fitness Centers**	The number of fitness centers within 5 miles of each store.

Table 1: Summary of variables.

The **Weekly Sales** (*Units.Sold*) variable is a response variable and has to be explained by independent variables. At least the goal of the study is to check whether the variables **Sales Rep** (*Sales.Rep*) and **Endcap** (*Endcap*) significantly and positively affect the response variable.

Variable Name	Min	Max	No. of Unique Values
Units.Sold	47.56	1041.2	1386
Average.Retail.Price	2.89	6.25	582
Sales.Rep	0	1	2
Endcap	0	1	2
Demo	0	1	2
Demo1.3	0	1	2
Demo4.5	0	1	2
Natural	0	4	5
Fitness	0	5	6

Figure 1: Summary of Variables in the data set

Variables *Units.Sold* and *Average.Retail.Price* are quantitative variable while all other variables are dummy variables (see fig. 1). Dummy variables are discussed below.

1.1.1 Dummy Variables Impact

The data set provides seven **dummy** variables (see fig. 2). Only variables *Natural* and *Fitness* have more than two unique values while other dummy variables are binary. The table in fig. 2 provides intuition on how significantly dummy variables might affect *Units.Sold*. To access impact of dummy variables we compare **mean** and **standard deviation** of *Units.Sold* when each dummy variable is set to *zero* to **mean** and **standard deviation** of *Units.Sold* when each dummy variable is set to *non-zero*. We can see that variables *Sales.Rep*, *Endcap*, *Demo* and *Demo1.3* have substantial impact on *Units.Sold* **mean** and **standard deviation**. Meanwhile *Demo4.5*, *Natural* and *Fitness* seem to have no significant impact on *Units.Sold*.

Dummy Variable	No. of zeros (percentage)	Mean (sd) of Units.Sold when zero	No. of non–zeros (percentage)	Mean (sd) of Units.Sold when non–zero
Sales.Rep	624 (45%)	199 (58)	762 (55%)	299 (123)
Endcap	1333 (96%)	241 (74)	53 (4%)	584 (266)
Demo	1305 (94%)	245 (100)	81 (6%)	401 (163)
Demo1.3	1169 (84%)	235 (95)	217 (16%)	353 (137)
Demo4.5	1281 (92%)	248 (109)	105 (8%)	331 (106)
Natural	231 (17%)	229 (78)	1155 (83%)	259 (116)
Fitness	176 (13%)	267 (131)	1210 (87%)	252 (108)

Figure 2: Summary of dummy variables and impact of dummy variables on mean (standard deviation) of Units.Sold.

2 Methodology

2.1 Selecting Variables

We deployed All-Possible-Regressions Selection Procedure from **R** *leaps* package to select the best sets of variables for the linear model. We decided to chose between all linear first-order models with all possible interaction terms. For all tests we set *a priori* significance level of $\alpha = 5\%$.

Based on the procedure, we have identified **three** models with:

- minimal **BIC**;
- minimal **Mallow's Cp**;
- and maximal R^2_{adj}

respectively. The **R** *step* procedure from *MASS* package resulted in the best model with respect to minimization of **AIC** criterion. However, the **AIC**-best model was almost the same as **BIC**-best but also had insignificant regressors. Hence, we decided to discard **AIC**-best model in favor of **BIC**-best model.

	model	p - 1	SSE	R^2	R^2_a	AIC	BIC	PRESS	Cp
1	Min BIC Model	5	3439190	0.798	0.798	14781	14818	3469677	3.22
2	Min Cp Model	10	3409225	0.800	0.799	14779	14842	3468316	1.22
3	Max R^2_a Model	13	3399412	0.801	0.799	14781	14859	3475770	3.29

Figure 3: Comparison of three best MLR Models

As we can see (fig. 3), the three models share very similar R^2, R^2_{adj} and BIC while they contain five, ten and 13 independent variables respectively. As a matter of fact, when we compare the difference between R^2 and R^2_{adj} for the three models, it is easy to see that R^2 and R^2_{adj} are almost the same for model 1 (**BIC**-best model). As the number of regressors $(p-1)$ increases, R^2 increases in model 2 (**Cp**-best model) and model 3 (R^2_{adj}-best model) while R^2_{adj} decreases. This phenomenon indicates that potentially unnecessary regressors might be added to the model.

Further we check the summary for the 3 MLR models. We discovered that the coefficients are all significant in model 1, and the same variables stay significant in model 2 and 3 while other variables are less significant. The summary can be found in Appendix (see fig. 8).

As as result, we decided to choose the first model (see fig. 4) - the min BIC model is as follows:

$$Units.Sold = 189.59 + 106.29Demo + 75.98Demo1.3+$$

$$+73.1Demo4.5 + 52.05Sales.Rep + 461Sales.Rep : Endcap,$$

where the regressor $Sales.Rep : Endcap$ stands for interaction term of predictors $Sales.Rep$ and $Endcap$.

2.2 Refining the Model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	189.5865	2.0232	93.71	0.0000
Demo	106.2853	5.8020	18.32	0.0000
Demo1.3	75.9768	3.8139	19.92	0.0000
Demo4.5	73.0885	5.1236	14.27	0.0000
Sales.Rep	52.0502	2.8578	18.21	0.0000
Sales.Rep:Endcap	461.0027	8.5733	53.77	0.0000

Figure 4: Summary of Min BIC Model 1

We ran global F-test to test the significance of the independent variables as a group for predicting the response variable. The F-statistic is 1093 and the p-value is 2.2 10^{-16}, which is much smaller than $\alpha = 0.05$. Therefore we reject the null-hypothesis that all regression coefficients are zero in favor of alternate hypothesis that at least one regression coefficient is non-zero.

2.3 Testing Model Assumptions

2.3.1 Homoskedasticity of Residuals

We ran Breusch-Pagan Test to check homoscedasticity. The corresponding residual Plot can be found in **Appendix** (fig. 9). From the BP-test, we get a p-value of 0.2. For significance level $\alpha = 5\%$, there is no significant evidence to reject the null hypothesis that the residuals are homoskedastic.

2.3.2 Normality of Residuals

We created histogram and QQ-plot to visualize the distribution of residuals. Both histogram and QQ plot illustrate normal distribution of residuals. Histogram and QQ-plot can be found in **Appendix** (figs. 10, 11).

We also performed Shapiro-Wilk test to confirm the normality of residuals quantitatively. The p-value is 0.7. As a result, there is no significant evidence to reject null hypothesis that the residuals are normal.

2.3.3 Multicollinearity

As part of our analysis we addressed multicollinearity issue. The fig. 5 reveals no strong correlation between regressors chosen in model 1. However there is a very significant correlation between *Endcap* and the interaction term *Sales.Rep:Endcap*. This is the reason why we omitted *Endcap* regressor in model 1 even though the *Hierarchical Rule* demands to include first-order predictors from interaction term.

	Demo	Demo1.3	Demo4.5	Sales.Rep	Endcap	Sales.Rep:Endcap
Demo	1	0.04	–0.01	0.15	0.08	0.09
Demo1.3	0.04	1	0.01	0.24	0.09	0.1
Demo4.5	–0.01	0.01	1	0.14	0	–0.01
Sales.Rep	0.15	0.24	0.14	1	0.05	0.15
Endcap	0.08	0.09	0	0.05	1	0.82
Sales.Rep.Endcap	0.09	0.1	–0.01	0.15	0.82	1

Figure 5: Correlation Matrix

2.3.4 Independence of Residuals

However, we still cannot be completely sure that the model assumptions are satisfied. We also need to check residuals for auto-correlation (see fig. 6). This is essential for this case-study since the raw data was extracted from time series for various stores. Hence we can expect that the *true* errors might take the form

$$\epsilon_t = \rho\epsilon_{t-1} + u_t$$

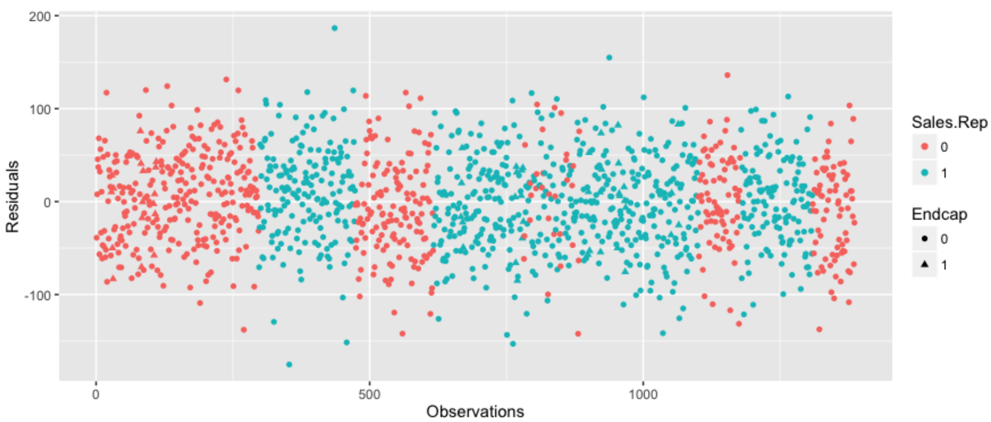

Figure 6: Residuals.

We ran Durbin-Watson to check residuals for serial correlation. The DW-statistic is 2.07 and the p-value is 0.276. Hence on $\alpha = 5\%$ significance level we cannot reject the null-hypothesis that residuals are not auto-correlated.

2.3.5 Outlying Cases

We used Cook's distance to measure the influences of observations. We define the threshold as $D_i > F(0.15, p, n-p)$. As shown below (see fig. 7), the red line represents the chosen threshold. In total there are **12** observations above the threshold. Thus we tried to remove those observations and rerun the model. The coefficients of the new model do not leave 95% confidence intervals of model 1. The new model summary and model 1's confidence interval can be found in **Appendix** (figs. 12, 13). As a result, we decided to continue with the original model 1.

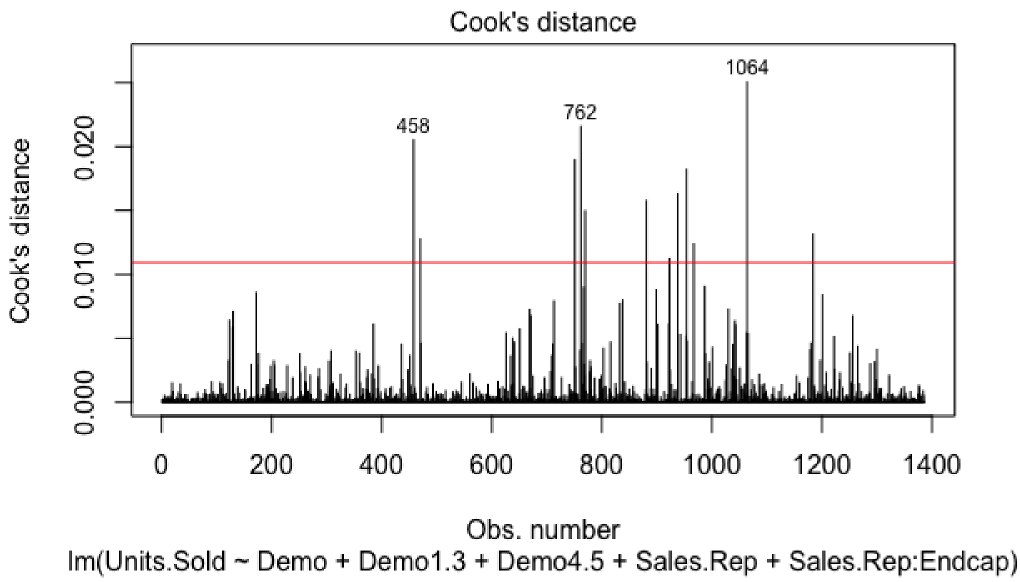

Figure 7: Cook's distance

3 Results

Based on the data provided we developed the model that:

- models **Weekly Sales (Volume)**;
- has minimal **BIC** criterion of all first-order models with interactions;
- has **positive** regression coefficients for **regional sales rep** and **endcap promotion** which are statistically significant on **5%** significance level;

- has residuals that are:
  - homoscedastic;
  - normally distributed;
  - and serially uncorrelated

  on **5%** significance level;

- and has no multicollinearity.

Based on this model we conclude that the impact of the in-store demos and endcap promotions is positive.

4 Conclusion

Basing on our model, in-store demos and endcap activities have positive impact on GoodBelly's sales. However, we would need more data to validate and cross-validate our model.

Still, the question of whether the company should proceed with promotions cannot be resolved based on the data set given. We showed that in-store demos and endcap promotions increase sales and how much they increase sales. However, the marginal increase from promotions has to be compared to the costs of such activities. It could also be the case that such promotions took place at stores that already have higher sales and managers are willing to participate in promotions because of relatively high sales.

It is worth noting that the best model takes into account the interaction term of presence of Sales Representative and endcap promotion. This regressor affect the sales level much higher than any other regressor (we can tell because all regressors are binary dummy variables). So we can conclude that the highest contribution to sales is from sales representatives who manage endcaps.

5 Appendix

	Dependent variable:		
	Units.Sold		
	BIC Model	Cp Model	R^2_a Model
	(1)	(2)	(3)
Demo	106.300***	101.600***	124.000***
	(5.802)	(6.494)	(16.360)
Demo1.3	75.980***	89.300***	89.640***
	(3.814)	(8.451)	(8.453)
Demo4.5	73.090***	66.800***	66.550***
	(5.124)	(14.140)	(14.140)
Sales.Rep	52.050***	54.770***	55.660***
	(2.858)	(3.067)	(3.124)
Demo:Demo1.3		24.820*	25.630*
		(14.300)	(14.800)
Demo:Natural			−7.265
			(6.264)
Demo:Sales.Rep			−17.510
			(16.320)
Demo:Endcap			28.810
			(21.270)
Demo1.3:Sales.Rep		−18.420*	−18.970**
		(9.551)	(9.566)
Demo4.5:Natural		7.579	7.854
		(5.043)	(5.046)
Demo4.5:Fitness		4.921	5.031
		(3.252)	(3.252)
Demo4.5:Sales.Rep		−22.140*	−23.230*
		(12.730)	(12.740)
Sales.Rep:Endcap	461.000***	460.200***	456.100***
	(8.573)	(8.571)	(9.351)
Constant	189.600***	188.400***	188.100***
	(2.023)	(2.099)	(2.117)
Observations	1,386	1,386	1,386
R^2	0.798	0.800	0.801
Adjusted R^2	0.798	0.799	0.799
Residual Std. Error	49.920 (df = 1380)	49.790 (df = 1375)	49.780 (df = 1372)
F Statistic	1,093.000*** (df = 5; 1380)	550.700*** (df = 10; 1375)	424.200*** (df = 13; 1372)
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure 8: Comparison of 3 MLR Models

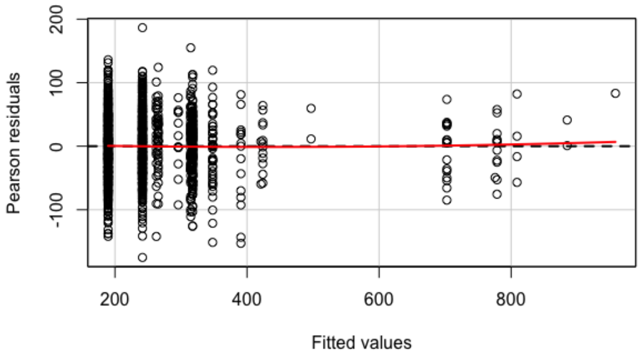

Figure 9: Residual Plot of Model 1

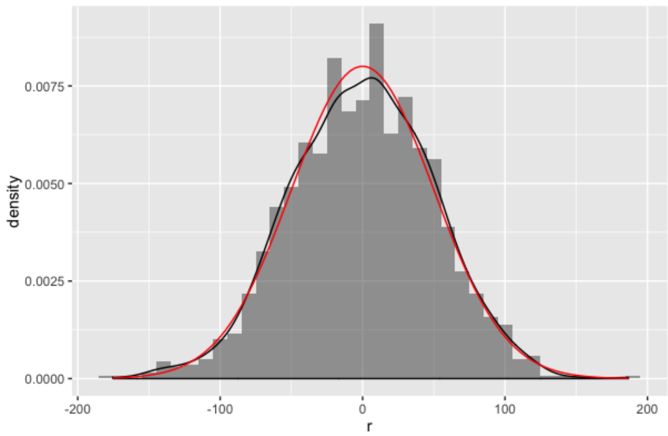

Figure 10: Histogram of Residuals of Model 1

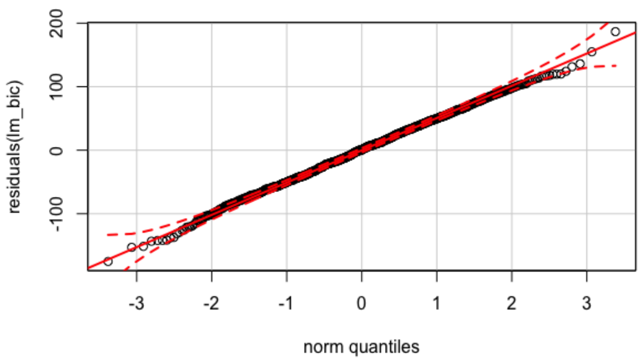

Figure 11: QQplot of Residuals of Model 1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	189.6679	1.9822	95.69	0.0000
Demo	105.9914	5.8388	18.15	0.0000
Demo1.3	77.4688	3.7632	20.59	0.0000
Demo4.5	74.9807	5.1342	14.60	0.0000
Sales.Rep	51.8572	2.8034	18.50	0.0000
Sales.Rep:Endcap	457.9916	9.0015	50.88	0.0000

Figure 12: New Model without influential points

	2.5 %	97.5 %
(Intercept)	185.62	193.56
Demo	94.90	117.67
Demo1.3	68.50	83.46
Demo4.5	63.04	83.14
Sales.Rep	46.44	57.66
Sales.Rep:Endcap	444.18	477.82

Figure 13: Confidence Interval of Coefficients of Model 1