Understanding promotional activity for beer products sold by supermarket chain

Retail Marketing Analytics - Individual Report

Julian West — CID 01623906

April 12, 2020

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vestibulum pretium libero non odio tincidunt semper. Vivamus sollicitudin egestas mattis. Sed vitae risus vel ex tincidunt molestie nec vel leo. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Maecenas quis massa tincidunt, faucibus magna non, fringilla sapien. In ullamcorper justo a scelerisque egestas. Ut maximus, elit a rutrum viverra, lectus sapien varius est, vel tempor neque mi et augue. Fusce ornare venenatis nunc nec feugiat. Proin a enim mauris. Mauris dignissim vulputate erat, vitae cursus risus elementum at. Cras luctus pharetra congue. Aliquam id est dictum, finibus ligula sed, tempus arcu.

Contents

1 Introduction

Retailers are interested in understanding the financial impact of promotional activity on product sales. The SCAN*PRO model is a popular method for quantifying the interaction effects between various promotions and allows retail store managers to improve their promotional mix decisions. In this analysis, the SCAN*PRO model will be used to understand the effects of promotional activity on the sales of three different brands of beer for a Chicago based supermarket store, Dominick's. The insights gained from the model will be used to make recommendations to the store manager on how to effectively use their promotional activity resource to maximise beer sales.

1.1 Business Objective & Research Question

The following business objectives and corresponding research questions were evaluated in the analysis:

Business Objective: What is the most effective form of promotion to boost unit sales?

- Which types of promotion are more effective for each beer brand, display or feature?
- Are promotions equally effective for each brand of beer?
- What is the price elasticity for each brand of beer?

This analysis will look at promotional activity from the perspective of the store manager. It is assumed that the manager has full control over feature and display promotions as well as setting the retail price of each beer brand. It is also assumed that there are no specific contractual obligations with the brand supplier guaranteeing a certain amount of promotion activity for their particular product.

The business objective aims to understand how the store manager can make best use of limited promotional resources. It is assumed that there is a finite amount of promotional resource and all products cannot be featured at the same time or all stock keeping units displayed at the same time. Therefore there is an benefit to knowing which brands perform better when on promotion so that the limited promotional resource can be allocated more effectively. It is also important to understand the price elasticity for each brand that pricing decisions can be made intelligently to increase the number of units sold. This may be important for clearing old stock at an optimum price.

1.2 About SCAN*PRO

SCAN*PRO is a store-level model which is able to quantify the effect of price changes and promotional activities on the unit sales of products for retailers. The model is a multiplicative model which includes own and cross brands for price, feature/display advertising, seasonality and a random error term (Leeflang et al. 2002):

$$q_{kjt} = \left[\prod_{r=1}^{n} \left(\frac{p_{krt}}{\bar{p}_{kr}} \right)^{\beta rj} \prod_{t=1}^{T} \gamma_{lrj}^{D_{lkrt}} \right] \left[\prod_{t=1}^{T} \delta_{jt}^{X_t} \right] \left[\prod_{k=1}^{K} \lambda_{kj}^{Z_k} \right] e^{u_{kjt}} \tag{1}$$

For $k = 1, ..., K, t = 1, ...T$

Where:

- q_{kjt} is unit sales for brand j in store k in week t
- p_{krt} is the unit price for brand r in store k in week t
- D_{1krt} is an indicator variable for display advertising: 1 if brand r is displayed 0 if not
- D_{2krt} is an indicator variable for feature advertising: 1 if brand r is featured 0 if not
- X_t is an indicator variable (for seasonal effects)
- Z_k is an indicator variable for store k
- e is an error term for unobserved or missing variables from the model

The parameters β_{rj}, γ_{lrj}. δ_{jt} and λ_{kj} are to be estimated. The β_{rj} parameters are the price elasticity's of the own brand and competing brands....

These parameters can be estimated using non-linear optimisation techniques, however, this can be computationally expensive, highly sensitive to model inputs and has no guarantee of reaching a global minimum. Therefore, it is easier to take the log transformation of equation 1 to make the model additive which can be solved using simple linear estimation techniques such as OLS.

The additive model thus becomes:

$$ln(q_{kjt}) = \sum_{r=1}^{n} \beta_{rj} ln \left(\frac{p_{krt}}{\bar{p}_{kr}} \right) + \sum_{r=1}^{n} \sum_{l=1}^{2} ln(\gamma_{lrj}) D_{lkrt} + \sum_{t=1}^{T} ln(\delta_{jt}) X_t + \sum_{k=1}^{K} ln(\lambda_{kj} Z_k + e_{kjt} \tag{2}$$

The model parameters from equation 2 can then be converted back into the multiplicative form in equation 1 to estimate unit sales for the week.

1.3 Software Used

All analysis for this project was conducted using Python (programming language). Details of code and packages used are included in the supplementary information.

2 Data

The data set consists of scanner records for beer category products from a Dominick's Finer Foods store (Dominick's). Dominick's was a Chicago-area grocery store chain before being acquired by Safeway in 1998 and eventually closing in 2013. The original source of the dataset was Chicago Booth Kilts Center for Marketing and a description of the dataset is included in (Srinivasan et al. 2004). However, the data set used for analysis was an adapted dataset shared by the Imperial College Retail Marketing Analytics teaching team via Slack (*beer_data_chicago_Dominicks.xlsx*).

The data covers 227 weeks (4 years 5 months) of sales for three brands of beer between 1989 and 1993. The raw data contains the following attributes for each brand:

Table 1: Raw data set features

Variable Name	Description
`SALESBRAND*`	the average sales ($) for brand *
`PRICEBRAND*`	the retail price ($) for brand *
`display_brand*`	the percentage of SKUs† on display for brand *
`FEATUREBRAND*`	the percentage of SKUs featured in the store for brand *
`RETAILERMARGINBRAND*`	the average retailer margin (%) for brand *
`WHOLESALEPRICEBRAND*`	the average wholesale price ($) for brand *

† Stock keeping units

The display and feature attributes are expressed as the percentage of stock keeping units (SKUs) of each brand which were promoted in that week. It is not clear from the literature (Srinivasan et al. 2004) what 'display' and 'feature' promotions specifically involved at the store so it is assumed that for a display promotion the product was displayed prominently in the store and for a feature promotion the product was 'on offer'.

The retailer margin is not a factor in the SCAN*PRO model described in equation 1 and was therefore not considered in the model estimation.

The wholesale price is not directly used in the SCAN*PRO model as the retail price is used instead. However, retail prices can be subjected to the endogeneity problem as they may be set be the store manager in response to how well the the product is selling and so they are not completely independent of the target variable, unit sales. The wholesale price can be a useful alternative to retail price as an instrument variable if endogenity exists. Therefore, the wholesale price was kept in the data set for further investigation.

The raw data did not contain a *unit sales* variable which is the desired target variable for the SCAN*PRO model. Therefore, a unit sales variable, *UNITSBRAND**, was created for each brand by dividing the sales of the brand by its average retail price for each week.

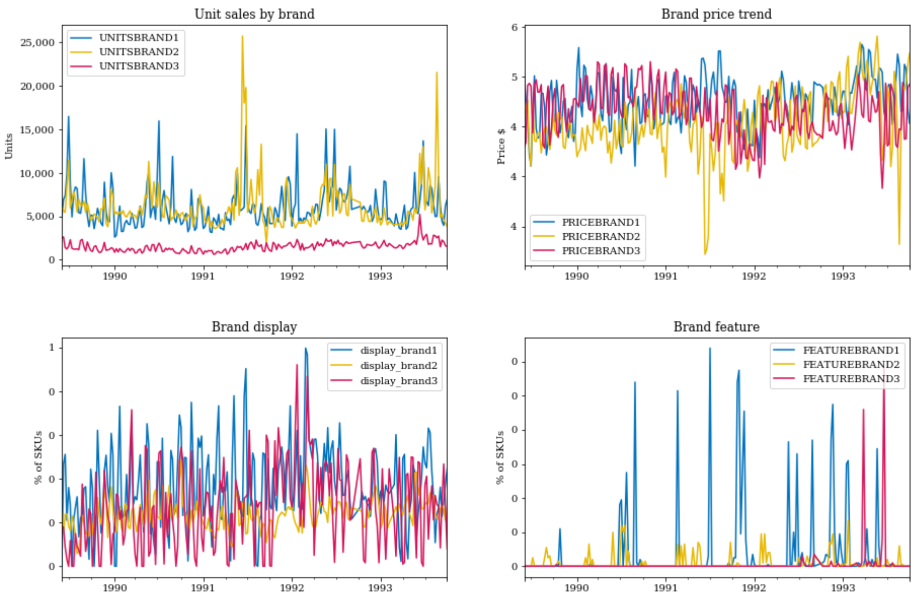

Figure 1: Sales, price, display and feature data for each brand

2.1 Assumptions

The data is given at a weekly level and will therefore trace the impact of price/display promotions over a weekly time period. It is assumed that the display and feature promotions lasted for the entire week.

The start date for the data observations was not explicit in the literature (Srinivasan et al. 2004), however, therefore it was assumed that the data observations started in the first week of June 1989 (4th June 1989). This assumption is consistent with the seasonality observed in figure 2.

2.2 Exploratory Analysis

The unit sales, prices, display and feature variables for each brand are shown in figure 1.

Brands 1 and 2 are the most popular with similar weekly unit sales but they also exhibit a high variation in unit sales from week to week. Whereas brand 3 has a much more consistent but lower sales volume. The retail sales price of each brand is similar and hovers between approximately $4 and $5 throughout the time period.

Each brand appears to have a certain percentage of stock keeping units 'on display' in almost every week. There is very high variation in the percentage of SKUs from week to week with brand 1 having the highest average percentage of SKUs on display (14%).

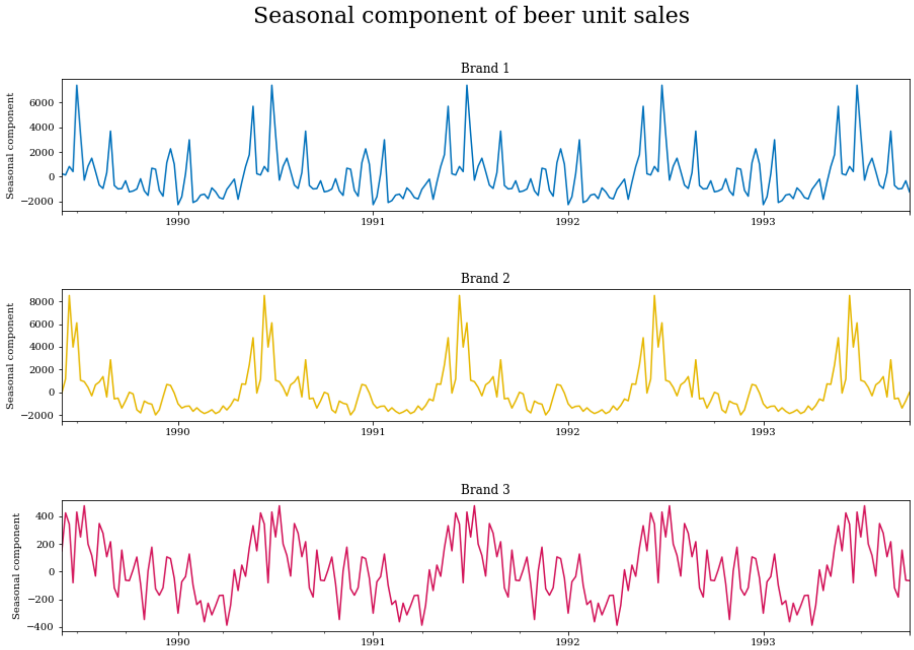

Figure 2: Seasonal component of unit sales for each brand unit sales

The 'feature' promotional activity appears to happen less frequently and used more sparingly than the 'display' promotions. Most weeks typically have no brands or just one brand featured'. Brand 1 had the most featured weeks and tended to have a greater percentage of SKUs featured in the week compared to when other brands were featured. Brand 2 was featured more consistently than the other brands, however, albeit with a smaller percentage of SKUs. Brand 3 was only featured towards the end of the time period of observations.

The SCAN*PRO model is also sensitive to seasonal effects in the unit sales. Figure 2 shows the seasonal components of the unit sales timeseries for each brand of beer. There are clear and significant seasonal trends in the data, with unit sales peaking in the summer months (May-August) as well as smaller peaks across the Christmas holiday period.

3 Model Estimation

3.1 Methodology

As there are three competing brands of beer, a separate model was created to estimate the unit sales of each brand. This yielded coefficients for own and cross-brand elasticises for price, display and feature promotions.

Table 2: R^2 for models trained on 80% of data and RMSE of model on unseen test data

	r2	rmse
Brand 1	0.792233	0.165940
Brand 2	0.812804	0.172020
Brand 3	0.813989	0.177375

3.1.1 Data Preparation

In order for the model to be estimated using linear methodologies (equation 2), the *UNITSBRAND**, *PRICEBRAND** and *display_brand** variables were log transformed. Due to the presence of a small number of zero values in the *display_brand** variables, a $log(x+1)$ transformation was used to avoid -inf values after the transformation.

The *FEATUREBRAND** variable could not be log transformed in this manner as the majority of values were zero for each brand. Therefore the *FEATUREBRAND** variable was encoded as a new binary variable, *binary_FEATUREBRAND**, which was 1 if the percentage of SKUs was greater than 0.02 and 0 otherwise. Even though the percentage of SKUs featured was not the same for each brand, the binary variable still preserved important information about the effects of feature promotions on unit sales.

Finally, eleven dummy variables $(k-1)$ were created for each month of the year to encode the seasonality observed in figure 2.

3.1.2 Model Estimation & Validation

The log unit sales (*log_UNITSBRAND**) for each brand were estimated using a regression model with the *log_PRICEBRAND**, *log_display_brand**, *binary_FEATUREBRAND** and seasonal dummy variables as inputs (see *01_regression_modelling.ipynb* in the supplementary information). [1]

To validate the models, 80% of the data was randomly sampled for model estimation (training) with 20% reserved for validating the accuracy of the model on unseen data. Once the model had been validated, the model was re-estimated using the full data set to obtain the final model coefficients which could be used for interpretation.

Table 2 shows the R^2 of the SCAN*PRO model for each brand trained on 80% of the available data. All three models gave reasonably good R^2 values with the models explaining approximately 80% of the variation in the data. The root mean squared error (RMSE) of the models on the unseen test data was approximately 0.17 for each model (equivalent to an average prediction error of $\pm 2\%$). The R^2 and RMSE for each model are consistent which suggests that the model is generalisable with good performance for all three brands.

[1] Alternative combinations of model inputs were also investigated, however, these were deemed to be less accurate and were not selected for further investigation (see *02_alternative_models.ipynb* in the supplementary information for analysis)

3.2 Results

The regression results for brand 1, brand 2 and brand 3 estimated using all available data observations are reported in table 3, 4 and 5 respectively.

Table 3: Brand 1: SCAN*PRO coefficients

Dep. Variable:	log_UNITSBRAND1	**R-squared:**	0.777
Model:	OLS	**Adj. R-squared:**	0.756
Method:	Least Squares	**F-statistic:**	35.96
Date:	Sat, 11 Apr 2020	**Prob (F-statistic):**	3.77e-56
Time:	16:22:19	**Log-Likelihood:**	94.246
No. Observations:	227	**AIC:**	-146.5
Df Residuals:	206	**BIC:**	-74.57
Df Model:	20		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	18.8007	1.033	18.206	0.000	16.765	20.837
log_PRICEBRAND1	-4.8740	0.369	-13.209	0.000	-5.601	-4.146
log_PRICEBRAND2	0.3262	0.205	1.589	0.114	-0.079	0.731
log_PRICEBRAND3	-1.4018	0.362	-3.873	0.000	-2.115	-0.688
log_display_brand1	0.6736	0.199	3.390	0.001	0.282	1.065
log_display_brand2	0.7605	0.414	1.837	0.068	-0.056	1.577
log_display_brand3	-0.3266	0.215	-1.521	0.130	-0.750	0.097
binary_FEATUREBRAND1	0.0939	0.040	2.376	0.018	0.016	0.172
binary_FEATUREBRAND2	0.0075	0.042	0.179	0.858	-0.075	0.090
binary_FEATUREBRAND3	0.0148	0.107	0.138	0.890	-0.196	0.225
August	0.2292	0.057	4.019	0.000	0.117	0.342
December	-0.0772	0.061	-1.266	0.207	-0.198	0.043
February	-0.1557	0.059	-2.625	0.009	-0.273	-0.039
January	-0.0700	0.060	-1.174	0.242	-0.188	0.048
July	0.2674	0.056	4.737	0.000	0.156	0.379
June	0.3694	0.059	6.304	0.000	0.254	0.485
March	-0.1076	0.060	-1.796	0.074	-0.226	0.011
May	0.3299	0.058	5.658	0.000	0.215	0.445
November	-0.0992	0.061	-1.633	0.104	-0.219	0.021
October	-0.0726	0.058	-1.244	0.215	-0.188	0.042
September	0.0766	0.055	1.383	0.168	-0.033	0.186

Omnibus:	11.282	**Durbin-Watson:**	1.446
Prob(Omnibus):	0.004	**Jarque-Bera (JB):**	11.511
Skew:	0.520	**Prob(JB):**	0.00317
Kurtosis:	3.366	**Cond. No.**	317.

Table 4: Brand 2: SCAN*PRO coefficients

Dep. Variable:	log_UNITSBRAND2	**R-squared:**	0.805
Model:	OLS	**Adj. R-squared:**	0.786
Method:	Least Squares	**F-statistic:**	42.52
Date:	Sat, 11 Apr 2020	**Prob (F-statistic):**	6.01e-62
Time:	16:36:52	**Log-Likelihood:**	111.22
No. Observations:	227	**AIC:**	-180.4
Df Residuals:	206	**BIC:**	-108.5
Df Model:	20		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	16.6887	0.958	17.416	0.000	14.799	18.578
log_PRICEBRAND1	-0.5023	0.342	-1.467	0.144	-1.177	0.173
log_PRICEBRAND2	-3.6822	0.190	-19.329	0.000	-4.058	-3.307
log_PRICEBRAND3	-0.6435	0.336	-1.916	0.057	-1.306	0.019
log_display_brand1	-0.1127	0.184	-0.611	0.542	-0.476	0.251
log_display_brand2	2.5825	0.384	6.724	0.000	1.825	3.340
log_display_brand3	-0.3336	0.199	-1.675	0.095	-0.726	0.059
binary_FEATUREBRAND1	0.0152	0.037	0.414	0.679	-0.057	0.087
binary_FEATUREBRAND2	0.0325	0.039	0.841	0.401	-0.044	0.109
binary_FEATUREBRAND3	0.0320	0.099	0.323	0.747	-0.163	0.227
August	0.1409	0.053	2.662	0.008	0.037	0.245
December	-0.0739	0.057	-1.306	0.193	-0.186	0.038
February	-0.1645	0.055	-2.989	0.003	-0.273	-0.056
January	-0.1237	0.055	-2.235	0.027	-0.233	-0.015
July	0.1566	0.052	2.989	0.003	0.053	0.260
June	0.2464	0.054	4.531	0.000	0.139	0.354
March	-0.0576	0.056	-1.035	0.302	-0.167	0.052
May	0.2229	0.054	4.121	0.000	0.116	0.330
November	-0.0993	0.056	-1.761	0.080	-0.211	0.012
October	-0.1148	0.054	-2.119	0.035	-0.222	-0.008
September	-0.0330	0.051	-0.642	0.522	-0.134	0.068

Omnibus:	8.722	**Durbin-Watson:**	1.414
Prob(Omnibus):	0.013	**Jarque-Bera (JB):**	16.292
Skew:	0.071	**Prob(JB):**	0.000290
Kurtosis:	4.305	**Cond. No.**	317.

Table 5: Brand 3: SCAN*PRO coefficients

Dep. Variable:	log_UNITSBRAND3	**R-squared:**	0.802
Model:	OLS	**Adj. R-squared:**	0.783
Method:	Least Squares	**F-statistic:**	41.69
Date:	Sat, 11 Apr 2020	**Prob (F-statistic):**	2.92e-61
Time:	16:37:01	**Log-Likelihood:**	100.73
No. Observations:	227	**AIC:**	-159.5
Df Residuals:	206	**BIC:**	-87.53
Df Model:	20		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	14.9740	1.004	14.920	0.000	12.995	16.953
log_PRICEBRAND1	0.4840	0.359	1.350	0.179	-0.223	1.191
log_PRICEBRAND2	1.0858	0.200	5.442	0.000	0.692	1.479
log_PRICEBRAND3	-6.1407	0.352	-17.455	0.000	-6.834	-5.447
log_display_brand1	0.1762	0.193	0.912	0.363	-0.205	0.557
log_display_brand2	1.5551	0.402	3.866	0.000	0.762	2.348
log_display_brand3	-0.4286	0.209	-2.055	0.041	-0.840	-0.017
binary_FEATUREBRAND1	0.0125	0.038	0.326	0.745	-0.063	0.088
binary_FEATUREBRAND2	-0.0133	0.040	-0.328	0.743	-0.093	0.067
binary_FEATUREBRAND3	0.1231	0.104	1.186	0.237	-0.082	0.328
August	0.2225	0.055	4.016	0.000	0.113	0.332
December	-0.0903	0.059	-1.522	0.129	-0.207	0.027
February	-0.1031	0.058	-1.789	0.075	-0.217	0.011
January	-0.0728	0.058	-1.256	0.210	-0.187	0.041
July	0.2690	0.055	4.903	0.000	0.161	0.377
June	0.2794	0.057	4.906	0.000	0.167	0.392
March	-0.1100	0.058	-1.888	0.060	-0.225	0.005
May	0.2266	0.057	3.998	0.000	0.115	0.338
November	-0.0538	0.059	-0.911	0.364	-0.170	0.063
October	-0.0464	0.057	-0.818	0.415	-0.158	0.065
September	0.1006	0.054	1.869	0.063	-0.006	0.207

Omnibus:	3.704	**Durbin-Watson:**	1.106
Prob(Omnibus):	0.157	**Jarque-Bera (JB):**	3.343
Skew:	0.244	**Prob(JB):**	0.188
Kurtosis:	3.339	**Cond. No.**	317.

4 Discussion

The R^2 values for each model range from 0.77 to 0.80, indicating that the models explain approximately 80% of the variation in log unit sales for each brand. For all three brands there are clear seasonal effects on The coefficients for each seasonal dummy variable. Sales of each brand are increased in the summer months with the dummy variables for May, June, July and August all having large and significant coefficients in all models.

The coefficients of the *binary_FEATUREBRAND**, *log_display_brand** and *log_PRICEBRAND** and their confidence intervals for each brand are shown in graphical form in figures 3, 4, 5.

4.1 Price

The price of the brand has the largest impact on unit sales. The own brand price elasticities of brand 1,2 and 3 are -4.874, -3.682, -6.141 respectively. Due to the log-log form of the model, these coefficients can be interpreted as meaning that a decrease in price by 1% will increase the unit sales by 4.874%, 3.682% and 6.141% for each brand respectively. This suggests that customers of brand 3 are most sensitive to changes in price. The error bar ranges for the own brand price elasticities are small compared to the magnitude of the coefficient which gives confidence of the sign and magnitude of these coefficients.

The cross-brand price elasticities are much lower than the own-brand price elasticities. Interestingly, reducing the price of brand 3 by 1% actually increases unit sales of brand 1 by 1.4%. Conversely, reducing the price of brand 2 by 1% reduces unit sales of brand 3 by 1.1%.

4.2 Display

The own brand display coefficients are 0.674, 2.583, -0.429 for each brand respectively. This implies that if the percentage of SKUs on display for each brand is increased by 1% (note this is not the same as an increase by 1 *percentage point*), then the unit sales for the week are increased by 0.674% and 2.583% for brands 1 and 2 whereas the unit sales are decreased by 0.429% for brand 3. Display promotions are therefore most effective for brand 2 with a significant uplift in sales for additional products on display. Display promotions are less effective for brand 1 and actually detrimental for brand 3. Brand 3 is the least popular of the three brands in terms of unit sales and customers may therefore not respond as well to the promotion.

4.3 Feature

The feature variables in the model are binary and their values are interpreted differently to price and display. The target variable was transformed using a $log(x+1)$ transformation, therefore the effect on unit sales of the *binary_FEATUREBRAND** variables is equal to $e^{coefficient}$ [REFERENCE]. The own brand feature coefficients for brand 1, 2

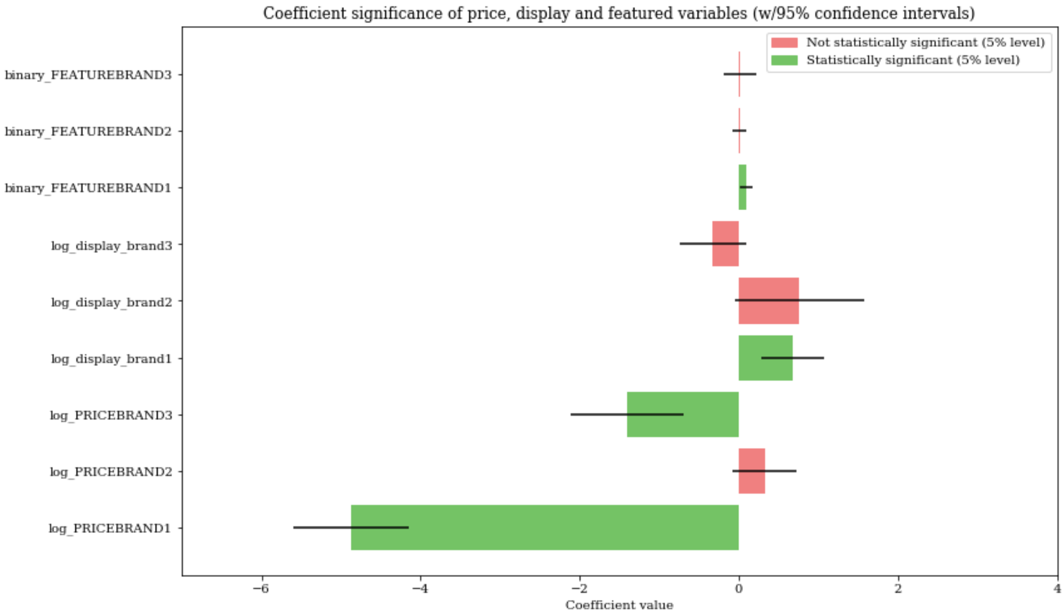

Figure 3: Brand 1 key model coefficients with confidence intervals

and 3 are 0.094, 0.033, 0.123 respectively. However, only brand 1's coefficient is statistically significant. Therefore, if brand 1 is featured during the week then unit sales are increased by approximately 9.9% ($e^0.094 = 1.099$). The lower 95% confidence interval for this coefficient is greater than 0, therefore it can be concluded that featuring brand 1 has a positive effect on sales.

The own brand coefficients for brand 2 and 3 are not statistically significant but are both positive. This suggests that featuring these brands in the week has a positive impact on sales as well, however, we can not be as confident in the magnitude of the effect.

5 Conclusions

5.1 Management Recommendations

The following conclusions and recommendations can be made to the store manager of Dominick's:

- Customers of brand 3 are most price sensitive
  - Increasing prices of brand 3 by 1% can reduce unit sales by 6.74%
- Display promotions are most effective for brand 2
  - Increasing the percentage of brand 2 SKUs on display by 1% increases unit sales by 2.58%

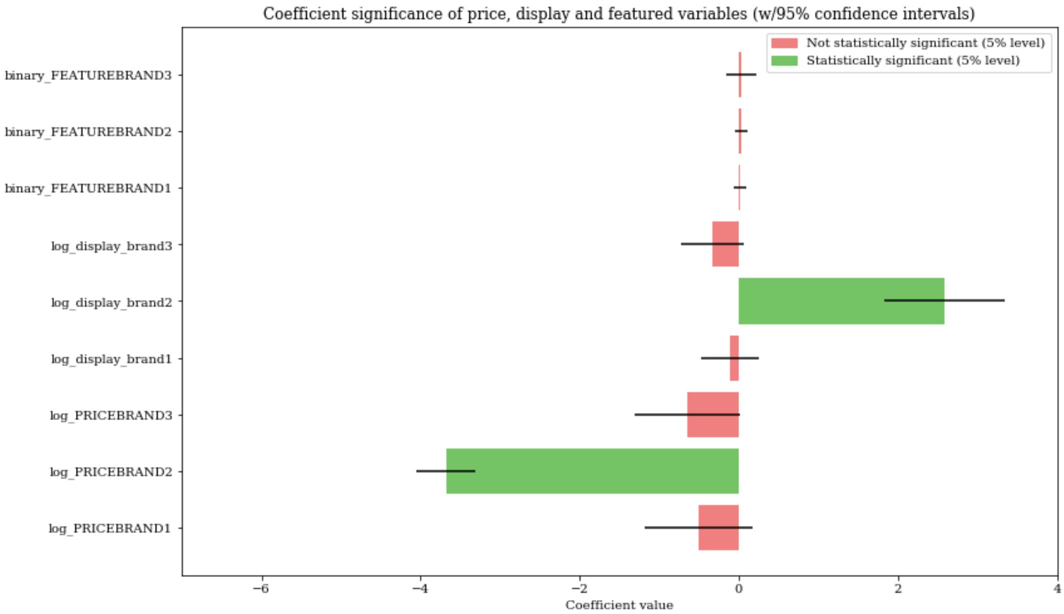

Figure 4: Brand 2 key model coefficients with confidence intervals

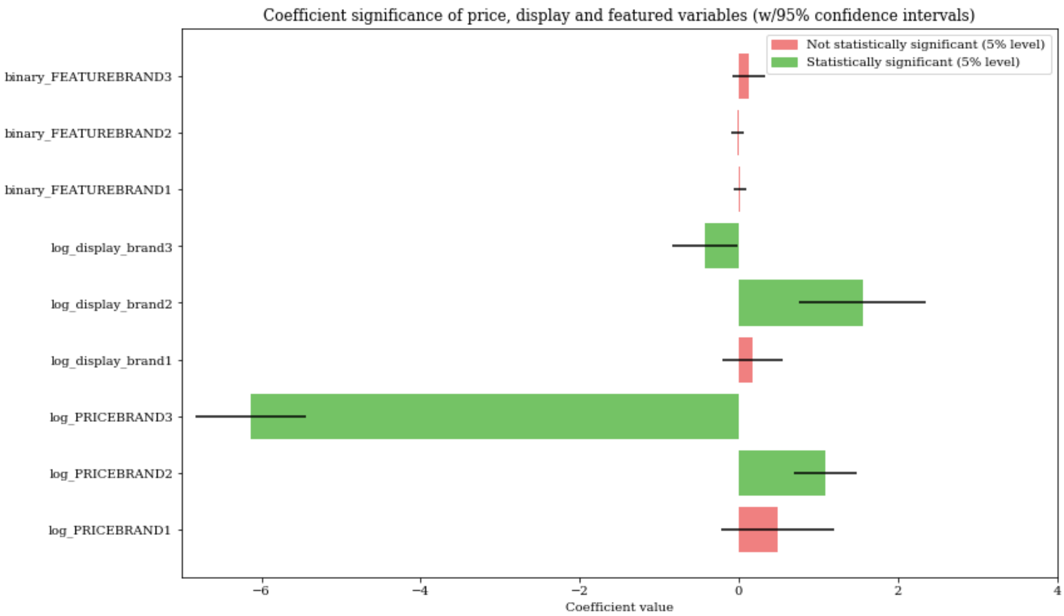

Figure 5: Brand 3 key model coefficients with confidence intervals

- Display advertising has only a minimal effect on unit sales for brands 1 and 3

- Feature promotions are most effective for brand 1
  - If brand 1 is on display in a given week, unit sales increase by 9.9%

5.2 Model Improvements

The ....

5.3 Future Work

The SCAN*PRO model only addresses the short-term impact of price changes and promotional activity on demand for the product. However, it would be useful for the store manager be able to understand the long-term impact of promotional activity on future unit sales. The delayed and forward purchase effects of promotions on customer purchasing behaviour can be quantified using VAR and IRFS models [REFERENCE] .

References

Leeflang, P. S., van Heerde, H. J. & Wittink, D. R. (2002), 'How promotions work: SCAN*PRO-based evolutionary model building'.
URL: *http://www.ssrn.com/abstract=321003*

Srinivasan, S., Pauwels, K., Hanssens, D. M. & Dekimpe, M. G. (2004), 'Do promotions benefit manufacturers, retailers, or both?', **50**(5), 617–629.
URL: *http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1040.0225*