Allocating Advertising Resources

University of California, Davis

Master of Science in Business Analytics (2020-21)

Section 1 Group D

Chetna Bhardwaj

Jessica Padolina

Jing Gan

Shreya Shah

Yuthika Agarwalla

Zixu “Theo” Zhou

Table of Contents

Executive Summary

It is vital for companies to understand how marketing and advertising affects sales to inform decisions regarding budget allocation. So it's worthwhile to study how each medium contributes towards sales generation and ensure the money is moving accordingly. In this analysis, a cosmetics firm has collected data for a product over four years, detailing the amount spent on each marketing channel and total sales generated monthly. The final framework is a dynamic one transformed by the square root function; meaning it captures the carryover effect of advertising and the economic theory of diminishing marginal returns. Not all variables were kept in the winning model to account for scarce data, statistical insignificance, and negative elasticity (hence, zero budget allocation). Post-analysis recommendations from the existing model include spending more on winback catalogs and portal advertising. Furthermore, there are other models included which better fit the dataset worth exploring and investing in.

Introduction

Advertisements are methods for companies to promote brands, attract new customers, retain existing ones, and increase profitability. The aim of this study is to identify the effects of ads on sales for a product launched four years ago by a cosmetics firm. It also aims to devise a multi-channel attribution model to optimally allocate advertising resources.

The firm employs a variety of offline and online advertising media to realize the company's integrated marketing communication objectives. This multi-media selection aims for simultaneous expansion and customer retention. Catalogs, direct mailing along with online media advertising have been the main media channels deployed to influence the target audience and trigger a response for purchase.

Problem Formulation

The goal is to create a model that goes beyond a simple correlation, closely captures media channels' cross-interactions, and measures elasticity to holistically predict future sales and allocate budget. Provided is data detailing ad spend and resulting sales for each month in a four-year period, via ten marketing mediums. The firm sends out different catalogs for existing customers, customers to win back, and new customers. The mail category encompasses flyers, postcards, and letters. Online platforms like banners, search ads, newsletters, portals, and social media are simultaneously employed.

The chosen framework details which are positively elastic and significantly impactful. Another phenomenon to consider is the carryover effect: ad dollars spent one month can influence sales volumes in consecutive months as customers take time to move through the purchasing funnel. This is included as a lag term transforming the static model into a dynamic one.

The focal model leverages square root function to portray diminishing returns, an economic theory that increasing factors of production after an optimal level actually result in lesser output. This mirrors the domain worldview because without intentional ad dollar allocation, the firm risks a "kitchen sink" of continuous spending, without significant effect on sales. The curve of the square root function is steeper near the origin and flattens further out, illustrating the aforementioned concept. Another transformation which captures this is the natural log function. Similar to the square root function, this graph flattens out as more resources are poured in, giving returns not directly proportionate.

Data Description

The provided data includes 42 months of sales and the ad expenditure over 10 media with

4 offline platforms and 6 online. On average, the offline media cost $935 per month, 9 times the total online advertising cost of $110. Such huge differences may be attributed to the nature of low marginal cost of online media, as well as lack of data on certain platforms.

Among the offline media, catalogs sent to the existing customers cost highest over all other media, on average $567 per month, taking over 50% of total monthly advertising because of its high monthly expenses and stable payment cycles. Catalogs sent to winback customers and new customers also cost much, $83 and $273 respectively, but displays less frequencies than catalogs sent to existing customers. On the other hand, online media advertising spending is more consistent and shows a rising trend over time. Among online media, searching ads cost highest with $70 monthly, and accelerated increasingly in the past 10 months. Newsletter spending and portal advertising have been relatively stable with minor increase for the latter. However, the data for banner ads and social media are mostly missing. We do not know whether those are included in the portal advertising package or simply not used by the firm, but decided to drop those two variables due to their incompleteness.

Model Development

Increased sales are not perfectly in step with advertising activities but rather delayed and spread out over a period of time. This 'carry-over effect' affecting the sales dynamics is expressed as a time lapse or lag structure (one month lagged sales). Our focal model is a simple autoregressive model with carryover effect and diminishing returns:

$Yt = \lambda Yt-1 + \beta 1\sqrt{X_{1t}} + \beta 2\sqrt{X_{2t}} + ... + intercept + \varepsilon_t$ where Yt is sales at time t, Xt are advertising efforts with diminishing returns at time t, intercept represents the mean level of initial sales in the absence of advertising, β_i is the effectiveness of advertising, λ is the carryover effect of advertising from the past sales $Y(t-1)$, and ε_t is an error term that represents the impact of other factors not explicitly included in the model. We are disregarding the confounding variables and endogeneity effect of advertising activities in this analysis.

The goal of model selection is to select the most parsimonious model supported by the observed advertising data with statistical significance and positive relation with sales. We have avoided variables which lead to negative coefficients in the model as the general requirement is positive elasticity between advertising mediums and sales. To balance parsimony and goodness-of-fit we have computed Akaike information criterion (AIC) and Schwarz's information criterion (BIC). We selected a model associated with the smallest values of the information criteria. Table 2 (Appendix) presents the AIC/BIC values for all models with and without parameters. Specifically, the proposed model attains the AIC value of 660.37 and BIC value of 668.97, which are the smallest values. Our final focal model as follow:

$$\text{Sales} = 2{,}095.36 + 0.19\ \text{Sales}\,(t-1) + 19.05\sqrt{CatalogsWinback} + 754.28\sqrt{Portals}$$

Results

We ran multiple regressions to derive the lowest AIC as our focal model and we calculated the elasticities for each advertising parameters. Advocate optimal advertising budget allocation is suggested to be proportional to advertising elasticities (Nerlove and Arrow, 1962).

From the results (Appendix C) coefficiently we shortlisted the Catalogs_Winback and Portals as the two main variables that affect the sales units with p = 0.009. Both variables have positive coefficient. The variance of Portals has the highest elasticity of 0.180 indicating that it is a major variance and deserves the most resource proportion. The variance of CatalogsWinback' elasticity is 0.025. The result is intuitively consistent to what we anticipated based on a business sense that 'Portals' is the main variance.

Recommendations and Managerial Implications

Management can benefit from exploring alternative models. We believe that there exist models that fit the dataset better than our focal model. Such alternatives are further explored in “Extensions”. Management should invest in these to minimize risks in misallocating budget to media based on an imperfect model. However, based on elasticity calculations, effectiveness estimates, and cost summary statistics, our recommendations are to firstly continue investing in winback catalogs and secondly continue or increase investing in portals.

Our study is incomplete without adequately exploring the effects of social media, banner ads, and retargeting. Today, brands use a common advertising strategy by promoting products through social media and influencers, who build relationships with consumers and stimulate brand loyalty (Qiutong 1) (see Appendix D). Thus, social media might affect sales more than catalogs or portals, and we strongly recommend that management prioritize collecting more data.

Conclusion

Our results show that the focal model is not the best-fit model for the relatively few sample data, and that the “log model with synergy” performs better. Resources should be allocated to catalogs, online search, and portals in this better-fit model. Management should focus on collecting more data on spends for social media, banner and retargeting advertisements, along with measuring the influence of factors such as seasonal trends and current brand positioning of the product in the market. Another aspect of analyzing advertising spends is to understand customer behavior. The company should look into the customer conversion funnel to track conversion at every point, thus identifying opportunities along the customer lifecycle for uptake and stickiness through effective advertising spends.

References

Simon, Julian L., Johan Arndt (1980). “The Shape of the Advertising Response Function”, J. Advertising Res. (20 August) 11–28.

Naik, Prasad A. and Raman, Kalyan (2003). “Understanding the Impact of Synergy in Multimedia Communications”, Journal of Marketing Research, Volume XL (Nov).

Nerlove, Marc and Kenneth Arrow (1962). “Optimal Advertising Policy Under Dynamic Conditions,” Economica, 29 (May).

Qiutong, Man; Rahman, Md. Jahidur (2019). “The Impact of Cosmetics Industry Social Media Marketing on Brand Loyalty: Evidence From Chinese College Students”, Academy of Marketing Studies Journal, Volume 23, Issue 2.

Appendix

TABLE 1: STATISTICAL SUMMARY FOR DATA

Variable name	**Description**	**Mea n**	**MIN. MAX.**
Sales (units)	Sales of items in units in the month	4809	3355 6976
ADV_Total	Total Advertising Spend in the month, comprises ADV_Offline and ADV_Online	1047. 16	59.61 1971.63
ADV_Offline	Total Offline Advertising Spend, comprises Catalogs_ Catalogs_NewCust in the month	935.3	0 1815.1
Catalogs_ExistC	Amount spent on Shopping Catalogs sent to existing	567.6	0

ust	Customers in the month		1298.7
Catelogs_Winback	Amount spent on Shopping Catalogs sent to Custome least 6 months) in the month	83.42	0 438.54
Catelogs_NewCust	Amount spent on Shopping Catalogs sent to New Customers in the month	272.87	0 1131.57
Mailings	Amount spent on Mailings (excluding Catalogs) sent flyers, postcards and letters in the month	11.42	0 84.47
ADV_Online	Total Online Advertising Spend, comprises Banner, S Retargeting and Portals in the month	111.84	50.41 295.21
Banner	Amount spent on Banner ads in the month	5.179	0 87.61
Search	Amount spent on Search ads in the month	69.83	38.17 134.87
SocialMedia	Amount spent on Social Media ads in the month	0	0 0
Newsletter	Amount spent on Newsletter ads in the month	20.73	7.057 53.61
Retargeting	Amount spent on Retargeting ads in the month	18.56	0 49.3
Portals	Amount spent on Portals ads in the month	5.25	2.544 9.303

TABLE 2: VALUES FOR INFORMATION CRITERIA

Models	**AIC**	**BIC**
Static model (without lagged term)	680.08	699.19
Focal model[1] (with lagged term)	665.65	686.22
Final Focal model	660.37	668.93

Final Focal Models contain diminishing returns and differential carryover. Model does not contain advertising medium - offline: Catalogs_ExistCust, Catalogs_NewCust, Mailings and online: Search, Banner, Social Media, Newsletter, Retargeting.

Regression output of Focal model[1] (with lagged term)

```
Call:
lm(formula = Sales ~ Stm1 + SqM1 + SqM2 + SqM3 + SqM4 + SqM5 +
    SqM6 + SqM8 + SqM9 + SqM10)

Residuals:
     Min       1Q   Median       3Q      Max
-1166.64  -481.27    45.89   368.91  1688.63

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1560.11816 1305.48737   1.195   0.2414
Stm1           0.07296    0.21486   0.340   0.7365
SqM1         -25.42388   17.45955  -1.456   0.1557
SqM2          49.71340   28.51097   1.744   0.0915 .
SqM3         -25.29200   15.65407  -1.616   0.1166
SqM4         -17.72565   46.26717  -0.383   0.7043
SqM5         147.32006  222.56314   0.662   0.5131
SqM6         -13.94797   62.94786  -0.222   0.8261
SqM8         125.69160  145.20000   0.866   0.3936
SqM9         -86.53916   95.53703  -0.906   0.3723
SqM10        866.98152  631.84484   1.372   0.1802
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 707.9 on 30 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.4065,    Adjusted R-squared:  0.2087
F-statistic: 2.055 on 10 and 30 DF,  p-value: 0.06231
```

Regression output of Final Focal Model:

```
Call:
lm(formula = Sales ~ Stm1 + SqM2 + SqM10)

Residuals:
    Min      1Q  Median      3Q     Max
-1278.3  -541.1  -114.7   404.2  1549.6

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2095.3639   811.3117   2.583   0.0139 *
Stm1           0.1943     0.1609   1.207   0.2350
SqM2          19.0518    16.1229   1.182   0.2449
SqM10        754.2801   294.4911   2.561   0.0146 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 708.9 on 37 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.2661,    Adjusted R-squared:  0.2066
F-statistic: 4.472 on 3 and 37 DF,  p-value: 0.008905
```

Appendix C: Elasticity Calculation

$$\text{Sales} = 2{,}095.36 + 0.19\ \text{Sales}\,(t-1) + 19.05\sqrt{CatalogsWinback} + 754.28\sqrt{Portals}$$

Name of Variance	**Elasticity Calculation**
$CatalogsWinback$	Elasticity of $\sqrt{CatalogsWinback}$ =1/2*19.05/(1-0.19)*sqrt(83.42) /4809=0.025
$Portals$	Elasticity of $\sqrt{Portals}$ =1/2*754.28/(1-0.19)*sqrt(5.246)/4809=0.180

Appendix D: The effect of social media marketing on brand loyalty in the cosmetics industry

A study on 145 college students in China conducted by Wenzhou-Kean University has shown that brand loyalty is positively related to social media content when campaigns include relevant and popular information, frequent updates, rewards mechanics, and variety in applications and platforms. The study was conducted using stepwise multiple regression, and brand loyalty was measured through college students' survey responses. The findings of this study support our recommendation that this cosmetics firm should explore the impact of social media ad spend before committing to an allocation model (see References).

Extensions

We ran robustness checks on several alternative models, which include:

1. Natural log of spend variable, dynamic
2. Natural log of spend variable, static
3. Square root focal model, no lag
4. Square root focal model, no intercept
5. Natural log, dynamic, dropping variables until performance was maximized
6. Natural log, dynamic, with synergy
7. Aggregated online and offline spend variable

In all tests, including the intercept and lagged sales term maximized model performance. While the lag term is never statistically significant, inclusion does improve model fit and is representative of real-life advertising inertia effects. As both terms have practical meaning for the firm, it is prudent to keep these variables in all future modeling attempts.

Revisiting the natural log model tested under *Problem Formulation*, we found that this model outperforms the square root focal model, and that performance maximized when 3 and 5 variables were included. We further explored advertising synergies as model predictors.

Advertising synergy is the combined effects of multiple media acting on brand awareness and sales. In a linear regression model, one can represent this using interaction variables between pairs of predictors. To select the best model, we prioritized parsimony and ease of explanation (this entailed eliminating negative coefficient terms where possible). One such better performing model is below:

$$Sales = 10{,}518.84 - 0.26 * Sales_{t-1} + 139.32 * [\ln(ExistCustCatalogs) + \ln(WinbackCatalogs)] + 230.08 * [\ln(WinbackCatalogs) + \ln(Newsletter)] + 922.47 * [\ln(Newsletter) + \ln(Portal)]$$

R output for this model is shown on the following page.

This model suggests positive effects on sales by the interaction between existing customer catalogs and winback catalogs. It is difficult to depict this interaction, as theoretically, existing customer catalogs and winback catalogs would not be sent to the same households. However, the other two interactions make more sense. It is realistic that winback catalogs and newsletters may both synergize to decrease the conversion rate among one-time or former customers. Similarly, one can imagine newsletter and portal ads having a similar synergistic effect on all customer types.

If the above model is to be followed, we repeat our recommendations of investing more money in missing online data collection, especially in social media. Additionally, we recommend continuing expenditure on existing and winback catalogs, but not new customer catalogs. Newsletter and portal ads should also receive more spending.

R output:

```
Call:
lm(formula = Sales ~ Stm1 + ln1 * ln2 + ln2 * ln8 + ln8 * ln10)

Residuals:
    Min      1Q  Median      3Q     Max 
-836.75 -324.28  -20.81  224.41 1048.71 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10518.8356  3120.8519   3.371  0.00197 ** 
Stm1           -0.2610     0.1353  -1.930  0.06252 .  
ln1          -307.8232    56.3770  -5.460 5.21e-06 ***
ln2         -1495.2616   313.3420  -4.772 3.85e-05 ***
ln8         -1428.8981   987.1541  -1.447  0.15749    
ln10        -1793.5969  1947.5710  -0.921  0.36397    
ln1:ln2       139.3168    53.4841   2.605  0.01383 *  
ln2:ln8       230.0782    96.5079   2.384  0.02322 *  
ln8:ln10      922.4741   650.3333   1.418  0.16572    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 480.7 on 32 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.7081,	Adjusted R-squared:  0.6352 
F-statistic: 9.704 on 8 and 32 DF,  p-value: 1.027e-06

> AIC(regmod18) #
[1] 632.5649
> BIC(regmod18)
[1] 649.7006
```