Multimedia Allocation

Student Names:

Anisha Samant

Sheetal Rajgure

Issam (Sam) Tamer

Siwei (David) Ran

School: *UC Davis*

Course Number and Name*: BAX 401 Information, Insight and Impact*

Instructor*: Ashwin Aravindakshan*

Date: *2022-10-27*

Table of Contents

Executive Summary: While designing their integrated media communications, firms need to allocate their budgets effectively across media. This is no longer proportional since multiple media interact with each other and their synergies contribute to the output. In this report, we analyze the effect of individual advertising media as well as the combined effect of their synergies on dependent variable (i.e. Sales) and arrive at a model that best fits the data based on AIC, BIC and Adj R squared scores – this analysis would be useful to effectively allocate advertising budgets to different media.

In our focal square root model, we identified the variables - "Catalogs_ExistCust", "Catalogs_Winback", "Catalogs_NewCust", "Search", "Newsletters", "Portals", the synergies between "Search" & "Portals", "Catalogs_ExistCust" & "Catalogs_Winback" and "Catalogs_ExistCust" & "Catalogs_NewCust" to be the best fit model. Further, we observed that the log model (natural log of advertising), with the lagged variable, captures the diminishing returns more efficiently. We calculated the elasticities to recommend change in budget allocations across media. Finally, we identified additional variables – granular social media data, seasonality, competitor's advertising spends and prices of ad inventories– that would help the model to predict and allocate more effectively.

Introduction: The advertising's impact on sales should no longer be measured one medium at a time. Different media interact with each other to influence a customer's decision. A Newsletter ad can prompt an internet search which could lead to clicking on a Google ad ending up in a sale. Understanding how the interaction takes place between different variables is essential to make sure budgets and ad spend are allocated in an optimal manner.

The cosmetics firm needs to know which advertising activities undertaken by them are effective in driving sales, and to quantify this effectiveness. The objective of this research paper is to analyze which and how the various advertising mediums of the cosmetics firm interact to collectively influence sales. The paper will discuss a chosen regression model to measure and attribute contribution of significant variables and thereby appropriately allocate resources to them.

Problem Formulation: The cosmetic firm uses various advertising activities spanning from online advertising like Social Media, Newsletter, Portals etc to offline activities such as Catalogs and Mailings for their product. Their data consists of their advertising spent across these mediums for the last 42 months. To select the best fit model, the appendix[4] of this paper consists of various combinations of linear models tested to see which provided the best AIC value.

Data Description: After careful consideration of all the variables in the dataset, we decided to exclude some non significant variables from our model. A detailed description of all the considered and selected variables are in appendix[1].

First, we eliminated the variables, “Banner” and “Social Media", which are composed majorly of null values throughout the dataset. Additionally, we excluded the variables “Retargeting” and “Mailings'' from our models; these variables had more than 50 percent null values in the dataset and they might make the results inaccurate because of the skewness.

Finally, our variables of interest used in our focal model include “Catalogs_ExistCust”, “Catalogs_Winback”, “Catalogs_NewCust”, “Search”, “Newsletters”, and “Portals”. We analyzed their correlation and five-number summary for the descriptive statistics in R and put it into appendix[2]. Although the “Catalogs_Winback” and “Catalogs_NewCust” variables also have missing data, their inclusion in our model, along with their synergies with other variables, which will be discussed in the next section, seem to improve the overall fit of our model based on the recorded AIC value.

Model Development: As we analyzed the problem and the dataset, we first created a model which regarded Sales as dependent on various advertisements. Considering the law of diminishing returns, the variables were transformed into functional forms like logs and square roots before running the models, and later compared to find the best fit. Assuming previous month’s advertising efforts could also be a significant variable in driving current sales, the models also accounted for this historic effect by creating lag variables so as to not over or underestimate the effects of the current month's ad spend.

$$Y_t(Sales) = \lambda Y_{t-1}(Past\ Sales) + \beta_{1t}\sqrt{Advertisement} + intercept + \epsilon_t$$

As we explained in the data description part, the advertisement in this dataset is made up of online sales and offline sales. Finally, we broke down online and offline advertisements to our chosen variables of interest. Taking into account that different advertising mediums will interact with each other to jointly influence sales, we added synergies between variables like budget spent on catalogs for existing customers and new customers, and between budget spent on catalogs for existing customers and customers who had not purchased in the past 6 months ("win-back"). The search and portal variables also had a significant interaction effect on the sales which we included in our model. These synergies capture the interaction effect between different variables. A change in one variable might not only affect the sales directly, but also change the impact of a different medium on sales. Synergies together give an overall better fitting model.

The synergies between variables were chosen in a way that improved our model's overall fit by comparing the AIC and adjusted R squared value.

$$Y_t(Sales) = \lambda Y_{t-1} + \beta_{1t}\sqrt{CataEx} + \beta_{2t}\sqrt{CataWin} + \beta_{3t}\sqrt{CataNew} + \beta_{4t}\sqrt{Search} +$$

$$\beta_{5t}\sqrt{NewsLetters} + \beta_{6t}\sqrt{Portals} + \beta_{7t}\sqrt{CataEx * CataNew} + \beta_{8t}\sqrt{CataEx * CataWin} +$$

$$\beta_{9t}\sqrt{Search * Portals} + \epsilon_t$$

Results: From our model, we can make observations about the impact that each advertising method has on sales. The coefficients and elasticities of significant variables of our focal model are listed in Appendix [3] From the limited data, we observe some unintuitive findings: the elasticities for catalogs (the only offline variables considered) are negative, meaning that as the amount spent on advertising catalogs increases, sales are expected to decrease. This implies that reducing advertising spent on catalogs would have a positive impact on sales. Online search advertising & Portals are more in line with expectations: across the board, an increase in online advertising leads to an increase in sales.

We can use these calculated elasticities to recommend a reallocation of the existing advertising budget according to the "Allocation" column. Here, the model leads us to a recommendation of increasing all of our online advertising (with portals seeing a striking 100% increase), while proportionally decreasing their offline advertising spending. However, as discussed in the next section, this recommendation is not founded on complete information.

Recommendations and Managerial Implications: The quality of data available greatly limits the impact of regression analysis and the scope of any potential recommendations. From our focal model, we might be tempted to draw conclusions such as drastically cutting funding from other advertising sources to fund the Portals advertising medium, but, despite the data supporting it, this would be an uninformed decision. Our monthly data shows that, while data exists for Portals in every month, the actual amount spent ($\overline{Portals} = 5.25$) is negligible compared to other advertising, such as catalogs for existing customers ($\overline{CustEx} = 567.6$). One potential explanation for this would be Portals advertising is a pay-per-impression: as the number of sales increase, customers will more often see the portal and be exposed to less-impactful, post-purchase advertising. This leads to a strong correlation between Portals advertising budget and sales, without much actual causality to learn from it. By having a better understanding of each of the data, some of these concerns could be addressed.

Putting aside concerns regarding the quality and type of the data present, we can draw conclusions on the best model that models diminishing returns and how certain types of advertising interact with each other. Based on numerous different models detailed in the appendix, we've concluded that transforming each advertising variable using a natural logarithm most effectively reduces the model's AIC, a measure that assesses how well the model fits its data. We've also concluded that including interaction terms between catalog spending on existing versus new customers, existing versus win-back customers, and online search vs portal advertising were the most effective. Conceptually, a personal recommendation is an extremely effective source of advertising. By increasing spending on advertising to existing customers, newer customers who interact with existing customers might receive personal recommendations from a friend. As seen in our model, a negative synergy exists between search advertising and portal advertising. This also makes sense: increasing the amount of online advertising would decrease the isolated effectiveness of other online advertising, such as from portals.

Conclusion: Allocating budgets to multiple media to optimize the sales is not straightforward because of the synergies amongst various media. Diminishing returns from advertising calls for square root or log transformations of the independent variables. From our analysis, we conclude that the natural log model captures the diminishing returns and fits better with the data [*4*] having lowest AIC & BIC, it explains 69% variation in Sales and has the highest Adjusted R Squared among the models with Intercept and model without intercept explains 99.2% variation – we retained the intercept to avoid modeling that sales would drop to zero without an advertising budget, and to avoid artificially fitting the model.

The regression analysis we performed on the variables reveals the correlation and does not always imply causation. Other limitations of the model are missing values and lack of additional important variables. Data on social media further divided into platforms (for instance, Facebook, Instagram, Twitter etc.) could help us to understand which platforms are relevant for the cosmetics category and which are not. There is huge scope to extend the analysis by including seasonality, competitor's advertising spends data and prices of ad inventories. Seasonality would impact consumer buying behavior, competitor's advertising allocation would affect our market share and ultimately sales and the prices of ad inventories would affect the advertising units bought (in the same budget) affecting Sales.

References & Notes

1. Nichols, W. (2017, March 24). *Advertising Analytics 2.0*. Harvard Business Review. https://hbr.org/2013/03/advertising-analytics-20
2. Jon RognerudJon Rognerud and Chaosmap work with Fortune 500 companies. (2015, July 1). *Digital Advertising Analytics 2.0 - A Primer*. Chaosmap Digital Advertising & Marketing Agency | Los Angeles, CA. Retrieved October 27, 2022, from https://chaosmap.com/blog/digital-advertising-analytics-primer/

Appendix

1. Description of Variables

Variable Name used in models	**Variable**	**Description**
t	Months	Time in Months
Y	Sales (units)	Sales of items in units in the month
-	ADV_Total	Total Advertising Spend in the month, comprises ADV_Offline and ADV_Online
-	ADV_Offline	Total Offline Advertising Spend, comprises Catalogs_ExistCust, Catalogs_Winback, Catalogs_NewCust in the month
M1	Catalogs_ExistCust	Amount spent on Shopping Catalogs sent to existing Customers in the month
M2	Catalogs_Winback	Amount spent on Shopping Catalogs sent to Customers (who have not bought for at least 6 months) in the month
M3	Catalogs_NewCust	Amount spent on Shopping Catalogs sent to New Customers in the month
M7	Mailings	Amount spent on Mailings (excluding Catalogs) sent to Customers. Mailing include flyers, postcards and letters in the month
-	ADV_online	Total Online Advertising Spend, comprises Banner, Search, SocialMedia, Newsletter, Retargeting and Portals in the month

-	Banner	Amount spent on Banner ads in the month
M4	Search	Amount spent on Search ads in the month
-	SocialMedia	Amount spent on Social Media ads in the month
M5	Newsletter	Amount spent on Newsletter ads in the month
M8	Retargeting	Amount spent on Retargeting ads in the month
M6	Portals	Amount spent on ad portal advertising in the month

1.1 Defining the variables:

(M1: Catalog_Existing M2: Catalog_New M3: Catalog_Winback M4: Search M5: Newsletter M6:Portals M7:Mailing M8:Retargeting)

[sqrt means Square root]
SqM1 = sqrt(Catalog_Existing)
SqM2 = sqrt(Catalog_New)
SqM3 = sqrt(Catalog_Winback)
SqM4 = sqrt(Search)
SqM5 = sqrt(Newsletter)
SqM6 = sqrt(Portals)
SqM7 = sqrt(Mailings)
SqM8 = sqrt(Retargeting)
[Log_e means natural log]
LogM1 = Log_e(Catalog_Existing)
LogM2 = Log_e(Catalog_New)
LogM3 = Log_e(Catalog_Winback)
LogM4 = Log_e(Search)
LogM5 = Log_e(Newsletter)
LogM6 = Log_e(Portals)
LogM7 = Log_e(Mailings)
LogM8 = Log_e(Retargeting)

2. Descriptive Statistics on variables of interest

a. Five Number Summary:

	Minimum	Q1	Median	Q3	Maximum
SqM1	0.000000	17.999182	24.453103	25.013112	36.037393
SqM2	0.000000	0.000000	4.670732	22.704384	33.638758
SqM3	0.000000	0.000000	0.000000	13.369460	20.941328
SqM4	6.177896	6.704591	8.129440	9.404120	11.613422
SqM5	2.656415	4.081362	4.447316	5.061017	7.321842
SqM6	1.595126	1.841893	2.169567	2.623000	3.050143

b. Correlation between variables of interest

	SqM1	SqM2	SqM3	SqM4	SqM5	SqM6
SqM1	1.00000000	0.01404391	0.04468071	-0.01880519	0.32042146	0.01073198
SqM2	0.01404391	1.00000000	0.75060019	0.01260171	0.04936839	0.03598879
SqM3	0.04468071	0.75060019	1.00000000	-0.14849819	-0.07827340	-0.05554444
SqM4	-0.01880519	0.01260171	-0.14849819	1.00000000	0.07896467	0.89030759
SqM5	0.32042146	0.04936839	-0.07827340	0.07896467	1.00000000	0.13780991
SqM6	0.01073198	0.03598879	-0.05554444	0.89030759	0.13780991	1.00000000

3. Elasticities

Variable	Coefficient (β)	Elasticity (η)	Pr(>|t|)	Allocation
Lagged Sales	-0.2599	N/A	0.190105	N/A
Sqrt(Catalog (Existing Customer))	-74.1141	-0.0290	0.000364 ***	-6.80%
Sqrt(Catalog (New Customer))	-66.5495	-0.0184	0.241331	-4.32%
Sqrt(Catalog (Win Back))	-108.9358	-0.0192	0.202037	-4.50%
Sqrt(Online Searches)	604.3295	0.1430	0.001966 **	33.54%
Sqrt(Newsletters)	79.2100	0.0138	0.466381	3.24%
Sqrt(Portals)	3470.5528	0.4298	0.000219 ***	100.82%
Interaction between catalog's impact of sqrt(existing) and sqrt(new) customers	2.6079	0.0034	0.346285	N/A
Interaction between catalog's impact of sqrt(existing) and sqrt(win-back) customers	5.5836	0.0052	0.157232	N/A
Interaction between online sqrt(searches) and sqrt(portals)	-283.7519	-0.1023	0.000222 ***	N/A

4. Testing other model specifications for robustness

We explored different combinations of variables with natural log and square-root transformations. Detailed results and comparison with the selected best model are given below

Model1:

Square root Model without synergy with least AIC and highest Adj R_sq (with intercept)

$$Y_t = \lambda Y_{t-1} + \beta_1 SqM1_t + \beta_2 SqM2_t + \beta_3 SqM3 + \beta_4 SqM4_t + \beta_5 SqM5_t + \beta_6 SqM6 + Intercept + \epsilon_t$$

Model2:

Square root Model without synergy with least AIC and highest Adj R_sq (without intercept)

$$Y_t = \lambda Y_{t-1} + \beta_1 SqM1_t + \beta_2 SqM2_t + \beta_3 SqM3 + \beta_1 SqM4_t + \beta_2 SqM5_t + \beta_3 SqM6 + \epsilon_t$$

Model3:

Square root Model with synergy with intercept

$$Y_t = \lambda Y_{t-1} + \beta_1 SqM1_t + \beta_2 SqM2_t + \beta_3 SqM3 + \beta_4 SqM4_t + \beta_5 SqM5_t + \beta_6 SqM6$$

$$+ \beta_7 SqM1_t * SqM2_t + \beta_8 SqM1_t * SqM3_t + \beta_9 SqM6_t * SqM4_t + Intercept + \epsilon_t$$

Model4:

##Square root Model with synergy without intercept

$$Y_t = \lambda Y_{t-1} + \beta_1 SqM1_t + \beta_2 SqM2_t + \beta_3 SqM3 + \beta_4 SqM4_t + \beta_5 SqM5_t + \beta_6 SqM6$$

$$+ \beta_7 SqM1_t * SqM2_t + \beta_8 SqM1_t * SqM3_t + \beta_9 SqM6_t * SqM4_t + \epsilon_t$$

Model5:

Log Model without synergy with least AIC and highest Adj R_sq (with intercept)

$$Y_t = \lambda Y_{t-1} + \beta_1 LogM1_t + \beta_2 LogM2_t + \beta_3 LogM3 + \beta_4 LogM4_t + \beta_5 LogM5_t + \beta_6 LogM6$$
$$+ Intercept + \epsilon_t$$

Model6:

Log Model without synergy with least AIC and highest Adj R_sq (without intercept)

$$Y_t = \lambda Y_{t-1} + \beta_1 LogM1_t + \beta_2 LogM2_t + \beta_3 LogM3 + \beta_4 LogM4_t + \beta_5 LogM5_t + \beta_6 LogM6 + \epsilon_t$$

Model7:

Log Model with synergy with intercept

$$Y_t = \lambda Y_{t-1} + \beta_1 LogM1_t + \beta_2 LogM2_t + \beta_3 LogM3 + \beta_4 LogM4_t + \beta_5 LogM5_t + \beta_6 LogM6$$
$$+ \beta_7 LogM1_t * LogM2_t + \beta_8 LogM1_t * LogM3_t + \beta_9 LogM6_t * LogM4_t + intercept + \epsilon_t$$

Model8:

Log Model with synergy without intercept

$$Y_t = \lambda Y_{t-1} + \beta_1 LogM1_t + \beta_2 LogM2_t + \beta_3 LogM3 + \beta_4 LogM4_t + \beta_5 LogM5_t + \beta_6 LogM6$$
$$+ \beta_7 LogM1_t * LogM2_t + \beta_8 LogM1_t * LogM3_t + \beta_9 LogM6_t * LogM4_t + \epsilon_t$$

Model9:

Log Model without lagged sales (with intercept)

$$Y_t = \beta_1 LogM1_t + \beta_2 LogM2_t + \beta_3 LogM3 + \beta_4 LogM4_t + \beta_5 LogM5_t + \beta_6 LogM6$$

$$+ Intercept + \epsilon_t$$

Model10:

Log Model without lagged sales (without intercept)

$$Y_t = \beta_1 LogM1_t + \beta_2 LogM2_t + \beta_3 LogM3 + \beta_4 LogM4_t + \beta_5 LogM5_t + \beta_6 LogM6 + \epsilon_t$$

Model11:

##Square root Model without synergy without intercept another variables set (Search, Portals, Catalog_Existing and Retargeting)

$$Y_t = \lambda Y_{t-1} + \beta_1 SqM1_t + \beta_2 SqM4_t + \beta_3 SqM8_t + \beta_4 SqM6 + Intercept + \epsilon_t$$

Model12:

##Square root Model with synergy without intercept another variables set (Search, Portals, Catalog_Existing and Retargeting)

$$Y_t = \lambda Y_{t-1} + \beta_1 SqM1_t + \beta_2 SqM4_t + \beta_3 SqM8_t + \beta_4 SqM6 + \beta_5 SqM1_t * SqM4_t +$$

$$+ \beta_6 SqM8_t * SqM4_t + \beta_7 SqM6_t * SqM4_t + \beta_8 SqM6_t * SqM8_t + \epsilon_t$$

5. Results of different models:

Model No & transformation	No of significant βs*	Std Error	P-value: overall model	R_sq	Adj R_sq	AIC	BIC	Variables of interest
1 Sqrt	3	686	0.015	38%	26%	661	676	Catalog-New, Catalog-Ex, Catalog-Win, Search, Portals, and Newsletters
2 Sqrt	4	709	< 2.2e-16	98%	97.8%	662	676	
3 Sqrt	1	550	0.000161	64%	52%	645	665	
4 Sqrt	4	554	< 2.2e-16	99%	98.7%	643	662	
5 Log	4	593	0.000260	54%	45%	649	664	
6 Log	4	588	<2.2e-16	98.8%	95.5%	648	661	
7 Log	**1**	**511**	**2.247e-05**	**69%**	**59%**	**639**	**660**	
8 Log	4	507	<2.2e-16	99.2%	98.9%	638	656	
9 Log	3	582	3.017e-05	56.4%	49%	662	676	
10 Log	5	582	< 2.2e-16	98.8%	98.6%	661	674	
11 Sqrt	2	695	0.01151	33%	24%	660	672	Search, Portals, Catalog-Ex, and Retargeting
12 Sqrt	4	694	0.03407	41%	23.9%	663	682	

*(at 10% significance level)

Log model (#7) data:

Variable	**Coefficient (β)**	**Elasticity (η)**	**Pr(>|t|)**
(Intercept)	-6393.84	N/A	0.4220
ln(catalog existing)	-235.07	-39.96	0.000039***
ln(catalog new)	-26.40	-8.59	0.9660
ln(catalog win-back)	-1034.21	-518.08	0.1970
ln(search)	2446.46	582.95	0.2040
ln(newsletter)	182.22	60.58	0.4310
ln(portals)	6429.27	3602.41	0.1850
interaction between ln(catalog existing) and ln(catalog new)	-15.01	-0.79	0.8810
interaction between ln(catalog existing) and ln(catalog win-back)	198.44	16.02	0.1210
interaction between ln(search) and ln(portals)	-1282.18	-168.97	0.2450

6. R Code for selected model:

```
Sales<-Mult$`Sales (units)`
Catalogue_Existing<-Mult$Catalogs_ExistCust
Catalogue_New<- Mult$Catalogs_NewCust
Catalogue_Winback<- Mult$Catalogs_Winback
Search <- Mult$Search
Newsletter<- Mult$Newsletter
```

```
Portals <- Mult$Portals

LogM1 = log(Catalogue_Existing+1)
LogM2 = log(Catalogue_New+1)
LogM3 = log(Catalogue_Winback+1)
LogM4 = log(Search+1)
LogM5 = log(Newsletter+1)
LogM6 = log(Portals+1)

## Log Model with synergy with intercept
regmod4<-lm(Sales ~ LogM1 + LogM2 + LogM3 + LogM4 + LogM5 + LogM6 + LogM1*LogM2 +
LogM1*LogM3 + LogM6*LogM4)
summary(regmod4)
AIC(regmod4)
```