Marketing Analytics

Assignment 1 (50 credits)

In June 1996, Kraft Foods Australia was involved in a major product-harm crisis as it had to recall two of its biggest brands of peanut butter (Eta and Kraft) because of potential contamination with salmonella. Over 100 cases of salmonella poisoning were reported, and the media and health authorities attacked Kraft for responding slowly to the crisis. The distribution of Kraft was completely down for more than four months. In fall 1996, both Eta and Kraft were reintroduced in the shelves, a reintroduction that was accompanied by a major Kraft advertising campaign. Your job is to find out whether investing in advertising after the crisis was an effective strategy for Kraft Foods.[1] **Hand in a word document of maximum 15 pages (all included!) before the start of the class on March 3 (line spacing 1.5, font size 12).** Only show the output that is strictly necessary for your discussion. **The focus of this assignment should be on the interpretation! Show that you really understand the performed analyses. There is no need to test the assumptions of the linear regression.**

Assignment1.sav is a dataset containing weekly sales (in kilos) of Eta and Kraft during one year after the reintroduction of the brands for Woolworth, the leading retailer in New South Wales. One observation is one week for the Eta or Kraft brand. The following variables are included in the dataset:

- **Eta**: this is a dummy variable indicating whether the observation is for Eta (eta=1) or Kraft (eta=0).
- **Kilos**: total volume sold in kilos.
- **Advertising**: total advertising spending of the affected brand in AU$.
- **AV_PRICE**: average price in AU$ per kilo.
- **Weeknr**: weeknumber, goes from 1 (for the first week after the crisis) to 52 (for the last observed week after the crisis).
- **Fall**: dummy variable indicating whether the observed week falls in the fall season (=0 if the week falls in another season).
- **Winter**: dummy variable indicating whether the observed week falls in the winter season (=0 if the week falls in another season).
- **Spring**: dummy variable indicating whether the observed week falls in the spring season (=0 if the week falls in another season).
- **Summer**: dummy variable indicating whether the observed week falls in the summer season (=0 if the week falls in another season).
- **Adv_comp**: total advertising spending of the competitors in AU$.

Below, you will find a description of the steps you need to take to assess the effectiveness of advertising spending after a product-harm crisis. We will build up the model step-by-step.

[1] For more information on the crisis case under study, please consult Van Heerde et al. 2007 (Marketing Science) and Cleeren et al. 2008 (Journal of the Academy of Marketing Science).

Each model builds further on the previous model (so just add the specified variables to the model you already had in the previous step) unless explicitly stated otherwise. In all steps, you need to run a model on Kraft and Eta together to assess the overall impact of after-crisis advertising, as well as a separate model for Kraft and Eta to assess potential inter-brand differences. In all analyses, you should just use the simple linear regression analysis with listwise deletion in case of missing values. Testing the assumption underlying linear regressions is not required.

1. Getting to know the data (3 credits)

A first step in building a market response model is always to get to know the data. Therefore, it is crucial to first do some exploratory data analyses.

a. Provide summary statistics for all variables in the dataset (look at average, standard deviation, minimum and maximum values, number of observations). Are there any particularities in the data? Explain your answer.
b. Another useful data inspection tool is the scatter plot. Plot sales and advertising spending for Eta and Kraft. Based on a visual inspection, would you expect a relationship between volume sold and advertising? Do you expect any inter-brand differences?

2. Advertising elasticities (7 credits)

Elasticities are often used to express the effectiveness of marketing instruments as they express by what constant percentage the dependent variable changes in response to a one-percent change in the marketing instrument. As a result, elasticities are unit free and can thus easily be compared throughout different contexts.

a. Define your *output and input measure*.
b. Estimate and interpret the *advertising elasticity* using the output and input measure for Kraft and Eta in a simple linear regression model. Write out the estimated model in an equation (use the equation editor in Word!). Discuss the fit of the model. Interpret the advertising elasticity. TIP: use the Compute variable (under the transform tab in SPSS) to create the natural logarithm (or ln) of a variable. To avoid too many missing values (because the log of zero is undefined), first add a small amount (e.g. 5) to all advertising expenditures before you take the natural logarithm.
c. Estimate an advertising elasticity *for Eta and Kraft separately*. TIP: you can use a selection variable when performing a linear regression in SPSS. Are there any differences?
d. What is your conclusion on the effectiveness of advertising spending in an after-crisis situation so far?

3. Adding control variables (8 credits)

To improve the model, and to enhance the estimated elasticities, environmental factors that could affect the output measure are added to the model.

a. Given that the dataset under consideration is a time series, we need to control for a number of typical characteristics of a time series dataset. As such, add the *long-term trend* to your model, and also control for the *season of the year*. Moreover, the *average price level* could also influence sales, and could hence be included in the model as an extra control variable. Write out this new model in an equation.
b. Estimate the three new models. Discuss the fit of the models. Does this change the estimated elasticities (either in total or separately for Kraft and Eta) and your conclusion on the effectiveness of advertising spending?
c. How do you interpret the estimated coefficients for the control variables? Discuss the face validity, i.e., do the coefficients make sense?
d. Which other environmental factors could potentially influence your output measure?

4. Adding the competition (10 credits)

Apart from the affected brand in a crisis, often also the non-affected competitors increase their advertising spending in an attempt to profit from the affected brands' misfortune. Therefore, we need to control for competitive advertising.

a. Add the natural log of competitive advertising to the model (add again 5 to all competitive advertising values before you take the log to avoid too many missing values). Write out the new equation. Discuss the fit of the models. Discuss the new estimated advertising elasticities (in total and for Eta and Kraft separately).
b. Interpret the coefficients of (log of) competitive advertising. What is your conclusion on the effectiveness of competitive advertising?
c. Another way to add competitive advertising to the market response model is to express the affected brand's advertising as a share of voice, i.e., $\frac{Advertising\ affected\ brand}{Advertising\ affected\ brand + Competitive\ advertising}$. Estimate the three market response models (total and for Eta and Kraft separately) using this share of voice expression rather than the model estimated under 4a (you can add the small value of 5 to the denominator in order to avoid a zero denominator). Note that taking the log of this Share Of Voice variable is not necessary given that it is already expressed in a unit-free way. Write out the new equation. Discuss the fit of the models and the new advertising elasticities. Does this change your conclusion on the advertising effectiveness?
d. Discuss the advantages and disadvantages of both methods to include competitive advertising in the market response model. Which method would you prefer?

5. Adding dynamics (18 credits)

It is unlikely that the payoff of an investment in advertising is limited to the same week. Therefore, we need to add some dynamics to our model. In order to keep the model simple, we will work on the basis of the model in step 3 (so without adding advertising of the competition).

a. One way to add dynamics to a market response model is to add leads and/or lags to the model. Discuss whether adding leads and lags to the advertising expenditures would make sense. Why?
b. Based on your discussion under a, complete your models with leads and/or lags with a bottom-up approach. Start your analysis with either one of both sources of dynamics (either leads or lags). Start with adding the first lead/lag to the model. If this coefficient is significant, add the second lead/lag etc. Continue this process until you find a model in which the last added lead/lag is insignificant. Your final model is then the previous model estimated (in which the last lead/lag is significant or no lead/lag if the first lead/lag is insignificant). Once this is completed, include the other source of dynamics (either leads or lags) to the final model established if you deem this necessary from your discussion in 5a. Do this analysis for Kraft and Eta together and for Eta and Kraft separately. Creating the lags or leads is easy by just copy pasting the right advertising values to a new variable in the SPSS data file (TIP: a good check to see whether you created the correct lead or lag is to check for which weeknumbers you have missing values). Alternatively, you can use the LAG function in the Compute Variable function of SPSS (make sure to also check the missing values here!). Show the output of your final models. Write out the final equations for the three models (total, Kraft and Eta).
c. Discuss the fit of your final models and discuss the effectiveness of advertising spending.
d. Another way to deal with dynamics in advertising spending is to work with an Adstock variable, i.e. $Adstock_t = \alpha \ln(Advertising_t) + (1-\alpha)\, Adstock_{t-1}$. Discuss how the Adstock variable works. What is the interpretation of α? How would you interpret an α value of 0 and 1? Which source of dynamics does it capture? Write out the new model equations.
e. Use a grid search to establish the optimal value for α in the three models (total, Eta and Kraft). Basically, this means that you run the model for different specifications of the Adstock variables (with different alpha values), and pick the model (and Adstock specification) that provides the best fit of the model (you can use R squared values as a fit measure). Try alpha values ranging from 0 to 1 using steps of 0.1. Given that Kraft and Eta were not distributed or advertised during the crisis, we can assume that $Adstock_0$ = 0 at the start of the crisis. TIP: you can easily create the different Adstock specifications in excel, and then copy paste them into your SPSS data file. Show the output of the three models. What do the acquired alpha values mean?

f. Discuss the fit of the three final models and the estimated Adstock coefficients.
g. Discuss the advantages and disadvantages of using the leads/lags and Adstock approach. Which of both methods do you prefer? Why?

6. Critical reflection and conclusion (4 credits)

As a final step in building a market response model, you should always think back on each modeling step that you took and reflect critically on the result.

a. Did every step really contribute to a better estimation of the advertising elasticities? In general, we like to stick to the rule "simpler is better". Would you advise against certain of the modeling steps above? Why?
b. Think about model validation and model testing? How could this be done?
c. What is your general conclusion on the effectiveness of advertising spending after a crisis? What would you advise companies to do?
d. Discuss the usability of these results in practice. If you think the results are not really usable, discuss ways to improve on the usability of the model.