Group Assignment 3

Advanced Econometrics

Etienne Wijler (Coordinator and Lecturer)

Noah Stegehuis (Coding instructor)

Gabriele Mingoli (Tutorial instructor)

Notes and instructions:

1. This assignment is mandatory.
2. The assignment is to be made in groups of 4 students. You can create your own group in Canvas > People > Assignment Groups, and then self-enroll or join an incomplete team. Please be professional and welcoming to new team members.
3. Only one of you needs to hand in all files.
4. The deadline for delivery of this assignment is on Friday, October 25, at 23:59h

   There will be no tolerance period for late deliveries. Deliveries after the assigned deadline imply that you have a final grade of zero for the assignment (GA3 = 0).
5. To get the full score for this assignment, the following three things must be done:
   (a) upload your final report as a PDF file in Canvas Assignments. Name the file **GA3report_2601842_2511351_2661510_2639486.pdf**, where the numbers are replaced by the VU student numbers of the 4 group members.To write you report according to academic standards follow the relevant tips that we have in the questions and also check the example report file under the name 'example_report.pdf' on Canvas > Assignments > Group Assignment 3: Instructions.

   (b) upload a zip file of your runnable R or Python code in Canvas Assignments. Name the file **GA3code_2601842_2511351_2661510_2639486_language.zip**, where the numbers are replaced by the VU student numbers of the 4 group members (or 3 if your group consists of 3 people) and language is either R or Python.

   The code file(s) should be clear, well commented, and directly runnable, so that it reads the datafile and obtains the results of all questions and prints them. Your initial comments in the file should hold your names and student numbers.

(c) upload a pdf of your entire code in Canvas Assignments. Name the file **GA3code_2601842_2511351_2661510_2639486_language.pdf**, where the numbers are replaced by the VU student numbers of the 4 group members (or 3 if your group consists of 3 people) and language is either R or Python. The file should be well readable, with proper indentations and should not contain pictures/photos/screenshots of code snippets.

6. As a standard anti-fraud measure, we will at random select a number of you to explain your code and answers. Any one of you must be able to explain any part of the code. Failure to explain your answers will result in a deduction of credits for this assignment.

7. For the support for the assignments, carefully read the announcement we put out at the start of the course and consult the discussion boards related to the assignments.

We wish you success!!

VU VRIJE UNIVERSITEIT AMSTERDAM | School of Business and Economics

1 Marketing and Sales: campaign optimization and price analysis

After having modelled volatilities for the financial risk management department and unemployment rates for the economics department, you are now ready for your final tasks at the Marketing & Sales department of Solara. You are informed that Solara has been an active player over the past ten years in the insurance market, offering products such as home insurance, contents insurance and travel insurance. The company has found that their insurance sales are quite sensitive to pricing and brand awareness. Up until now, however, key strategic decision related to dynamic pricing and marketing campaign budgets have typically been made solely on the basis of the experience of the marketing managers, without explicitly incorporating information from historical data. Accordingly, Solara wants you to analyze the effectiveness of their digital marketing campaigns, and to develop a first model for determining optimal product prices and marketing budgets in a data-driven manner. Given the complexity of this task and the breadth of their product portfolio, Solara has suggested you to focus solely on their single-trip travel insurances.

1.1 Marketing campaign effectiveness

Solara has historically allocated a fixed daily budget of 500 euro for digital marketing, spread evenly across Google Ads and Youtube. In order to understand the effect that variations in advertising may have on their revenue, Solara has implemented a small-scale AB-testing procedure for a period of 6 months. During this period, prices of their travel insurance were kept fixed, while random changes were made in their daily digital marketing expenditures. The resulting dataset can be found in the file `Solara_AB_data.csv`, which includes de-trended and de-seasonalized daily online sales (s_t), expenditures on advertisement directed at Google Ads (g_t), and expenditures on advertisement directed at Youtube (y_t), all measured in euros. During the AB-testing period, the daily add expenditures g_t and y_t were independently drawn from a $N(250, 80^2)$ distribution at each period. The sample of data shows that Solara spent about 88,155 euros over the 6 months of AB-testing, and achieved a total sales volume of around 635,609 euros over the same period of time.

In the field of marketing, the adstock created by an advertisement campaign refers to the prolonged effect that an advertisement campaign has on consumers. Adstock is created when consumers are exposed to advertisement, and it fades slowly over time since consumers will remember the advertisement for some time, also known as the decay effect. Increasing the amount of advertising increases the percentage of the audience reached by the advertising, but this relationship tends to be non-linear, with each incremental amount of advertising having a progressively lesser effect on demand. We will refer to the latter as the diminishing returns of advertising.

To take into account both the slow decay and diminishing returns, you consider the following model for the adstock generated by expenditure on Google and Youtube:

$$\text{gads}_t = \frac{\alpha_g g_t}{1+\delta_g \text{gads}_{t-1}} + \beta_g \text{gads}_{t-1}, \tag{1}$$

$$\text{yads}_t = \frac{\alpha_y y_t}{1+\delta_y \text{yads}_{t-1}} + \beta_y \text{yads}_{t-1}, \tag{2}$$

where gads_t (yads_t) denotes the adstock generated by the expenditure on Google search (Youtube) ads.

Finally, we consider the following linear models for the sales of Solara's single-trip travel insurances.

$$s_t = \mu + \phi_g \text{gads}_t + \phi_y \text{yads}_t + \epsilon_t. \tag{3}$$

Question 1. Explain the role of α_g, β_g and δ_g in the above Google Ads adstock model, as well as their connection (if any) to the described decay effect and the diminishing returns of advertising.

Question 2. Solara has informed you that based on previous research, they have found the following the estimates: $(\hat{\alpha}_g, \hat{\beta}_g, \hat{\delta}_g, \hat{\alpha}_y, \hat{\beta}_y, \hat{\delta}_y) = (2, 0.9, 10, 1, 0.97, 5)$. Filter the Google adstock and Youtube adstock using these parameter values. You may use $\widehat{\text{gads}}_1 = 22$ and $\widehat{\text{yads}}_1 = 41$ as initial values. Create a (2x2) grid of plots, where the left column contains plots of the filtered Google and Youtube adstocks, and the right column contains two histograms of the adstocks with 50 bins each. Describe and explain the patterns of the adstocks, paying particular attention to any skewness that may be present.

Question 3. Estimate the parameters of the following linear model

$$\text{s}_t = \mu + \phi_g \widehat{\text{gads}}_t + \phi_y \widehat{\text{yads}}_t + \epsilon_t.$$

using least squares regression. Interpret your estimates of ϕ_g and ϕ_y, assuming that the model is correctly specified.

Question 4. Show that the long-run adstocks that Solara would converge to, if the company would stick to their old strategy of assigning a daily fixed budget of

250 euro each to Google Ads and Youtube, is given by 22.31 and 40.72, respectively. Hint: taking Google Ads as an example, this question essentially asks to derive $\lim_{t\to\infty} \text{gads}_t$, where

$$\text{gads}_t = \frac{\alpha \cdot 250}{1+\delta \text{gads}_{t-1}} + \beta \text{gads}_{t-1}.$$

If the adstocks converge under a fixed investment strategy, this implies $|\text{gads}_t - \text{gads}_{t-1}| \to 0$ as $t \to \infty$. Accordingly, the limit adstock, say gads_∞ can be found by solving

$$\text{gads}_\infty = \frac{\alpha \cdot 250}{1+\delta \text{gads}_\infty} + \beta \text{gads}_\infty.$$

Question 5. Compute an Impulse Response Function that shows the effect of spending an extra 100 euro on Google Ads (on top of the daily fixed budget of 250 euros) on sales. For the Youtube adstock, you maintain the standard strategy of allocation a fixed daily budget of 250 euros. You may set the Google adstock prior to the impulse equal to $\text{gads}_j = 22.31$ (i.e. the long-term adstock from Question 4) and the sales equal to $s_j = \hat{\mu} + \hat{\phi}_g 22.31 + \hat{\phi}_y 40.72$ for all $j < t$, with t being the impulse date. Afterwards, perform the same analysis for Youtube (without providing the additional investment to Google).

Create a (2x2) grid of plots, with the first row corresponding to the impulse given to Google Ads and the second row corresponding to the impulse on Youtube advertisement. The left column should contain the daily responses of sales to the 100 euro impulse to Google Ads and Youtube, while the right column should contain the cumulative responses to the impulse. Each plot should contain 5 (constant) time points before the impulse, and 60 time points that include the effect of the impulse. Compare the differences in the responses of sales to impulses in Google Ads and Youtube.

Hint: using Google Ads as an example, this question requires computing gads_t with the parameters given in Question 2 and setting $g_t = 250 + 100$ at the impulse date. Also, note that $\mathbb{E}(g_{t+1} \mid \text{gads}_t) = 250$, since the expenditures were independently drawn during the AB-testing procedure.

Question 6. Is the AB-testing procedure important? Without AB-testing, would you be concerned about simultaneity issues and resulting endogeneity of regressors? How would that affect the interpretation of the results of your model? Explain your reasoning.

1.2 Dynamic Pricing

For the last project at the Marketing & Sales department, Solara asked you to help design a data-driven procedure for optimizing the prices of their single-trip travel insurance. To achieve this, Solara has provided you with the implied "costs" of their insurance (c_t), which they have calculated as the average declaration per insurance over a historical rolling window, combined with an additional markup to cover administrative and other costs. The data is de-trended and de-seasonalized, you do not have to worry about this. Furthermore, Solara has provided you with their historical marketing expenditures m_t. The resulting dataset can be found in the file `Solara_pricing_data.csv`.

Question 7. Plot the available data on sales, prices, acquisition costs and marketing expenditures.

Question 8. Run a simple regression of sales s_t on prices p_t,

$$s_t = \alpha + \beta p_t + \varepsilon_t \ . \tag{4}$$

Report the parameter estimates.

Question 9. Comment on the usefulness of this regression as a predictive model for sales. Suppose your estimator is consistent, to which limit will it converge? What are the properties of that limit?

Question 10. Comment on the value of the same regression as a structural model for analyzing the causal impact that a change in price p_t has on expected sales s_t. Explain your reasoning carefully.

Concerned by the possibility that price p_t is an endogenous regressor in (4) due to simultaneity, you decide to follow an instrumental variable approach.

Question 11. Run a regression of prices p_t on costs c_t,

$$p_t = \delta + \gamma c_t + u_t \ ,$$

and use predicted price $\hat{p}_t$ to find the structural-causal relation between sales and prices,

$$s_t = \alpha + \beta \hat{p}_t + \varepsilon_t \ .$$

Report the parameter estimates for both the instrumental regression and the sales regression. What is the causal impact of an increase in price p_t on expected sales s_t?

Question 12. Test for endogeneity of p_t in the original regression (4) using a Hausman-Durbin-Wu test. Do not use a built-in function or result in a summary table for the Hausman statistic, but code up the test yourself. If this calculation is missing in your code, zero points will be awarded for this question.

- Tip: For the Hausman-Durbin-Wu test you may use the following test statistic $H^T H = T(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})^T(\tilde{\boldsymbol{\Sigma}} - \hat{\boldsymbol{\Sigma}})^{-1}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})$, where $\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}$ are both consistent for $\boldsymbol{\theta}$ under H_0, but $\hat{\boldsymbol{\theta}}$ is an efficient estimator. Under the alternative $\tilde{\boldsymbol{\theta}}$ is consistent, while $\hat{\boldsymbol{\theta}}$ is not. Furthermore, $\sqrt{T}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \tilde{\boldsymbol{\Sigma}})$ and $\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \hat{\boldsymbol{\Sigma}})$ as $T \to \infty$. Under the null hypothesis the test statistic follows asymptotically the χ^2 distribution with k degrees of freedom, where k is the rank of $(\tilde{\boldsymbol{\Sigma}} - \hat{\boldsymbol{\Sigma}})$.

- Tip: For the OLS and 2SLS estimator in a linear regression model with data matrix $\boldsymbol{X}$ and instrument matrix $\boldsymbol{Z}$, we have asymptotic distributions
  $\sqrt{T}(\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}) \xrightarrow{d} N\left(\mathbf{0}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\right)$,
  $\sqrt{T}(\hat{\boldsymbol{\beta}}_{2SLS} - \boldsymbol{\beta}) \xrightarrow{d} N\left(\mathbf{0}, \sigma^2(\boldsymbol{Z}'\boldsymbol{X})^{-1}(\boldsymbol{Z}'\boldsymbol{Z})(\boldsymbol{X}'\boldsymbol{Z})^{-1}\right)$ as $T \to \infty$
  When estimating these variance matrices, carefully consider how to evaluate $\hat{\sigma}^2$ for each of the estimators. You may want to consult additional sources for this.

Question 13. Consider improving the sales model and the estimate of the causal relation between sales and prices by including marketing expenditures on the sales regression,

$$s_t = \alpha + \beta \hat{p}_t + \psi m_t + \varepsilon_t \ .$$

Report the parameter estimates you obtain.

Question 14. Using the model from Question 13 and last observed values in your sample, determine the optimal price to maximize the expected profit at time $T+1$. For each period t, the profit is defined

$$\pi_t = s_t(p_t - c_t) - m_t.$$

Hence, letting $\mathscr{F}_t = (s_t, p_t, c_t, m_t)$, your goal is to maximize $\mathbb{E}(\pi_{T+1} \mid \mathscr{F}_T)$ as a function of p_{T+1}. Since this optimization will involve the expressions $\mathbb{E}(c_{T+1} \mid \mathscr{F}_T)$ and $\mathbb{E}(m_{T+1} \mid \mathscr{F}_T)$, you may assume that

$$c_t = \mu_c + \phi_c c_{t-1} + \epsilon_{c,t}, \tag{5}$$

$$m_t = \mu_m + \phi_m m_{t-1} + \epsilon_{m,t}. \tag{6}$$

Estimating (5) and (6) then also allows you to compute $\mathbb{E}(c_{T+1} \mid \mathscr{F}_T)$ and $\mathbb{E}(m_{T+1} \mid \mathscr{F}_T)$, respectively. Report your derivations, your estimation results and your optimal price. Reflect on how this price compares to the last observed price P_T. Hint: since you are optimizing over P_{T+1}, the price should be interpreted as a parameter and not a random variable. This dramatically simplifies taking expectations. Use the slides on CH11 as a reference.