Group Assignment III
Advanced Econometrics

Emma Arussi, 2783830, e.s.arussi@student.vu.nl
Marta Chejduk, 2753970, m.m.chejduk@student.vu.nl
Niels van Herk, 2832099, n.a.a.van.herk@student.vu.nl
Tess Scholtus, 2831775, t.p.scholtus@student.vu.nl

Academic Year: 2024-2025

PART I: MARKETING CAMPAIGN EFFECTIVENESS

3.1

Parameter α
Parameter α represents the effectiveness of new advertising spending. It shows the immediate effect of spending on Google Ads and YouTube in the period t on the current Adstock level.

Parameter β
Parameter β represents adstock persistence or decay effect. This parameter captures the carry over effect (decay) of the previous period's adstock value. Specifically, it models how much of the adstock from the previous period persists into the current period.

Parameter δ
Parameter δ represents diminishing returns on advertising. The term in the denominator means that as the adstock from previous periods accumulates, it becomes harder for additional spending on Google Ads and YouTube in period to generate the same level of impact on sales. This captures the concept that additional advertising becomes less effective as a part of the audience has already been exposed to a significant amount of advertising.

Connection to Decay and Diminishing Returns

The decay effect is primarily captured by β which controls how quickly the impact of previous advertising fades over time.The diminishing returns are captured by δ which shows how much past advertising reduces the effectiveness of current advertising.

3.2

Figure 1 shows 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.

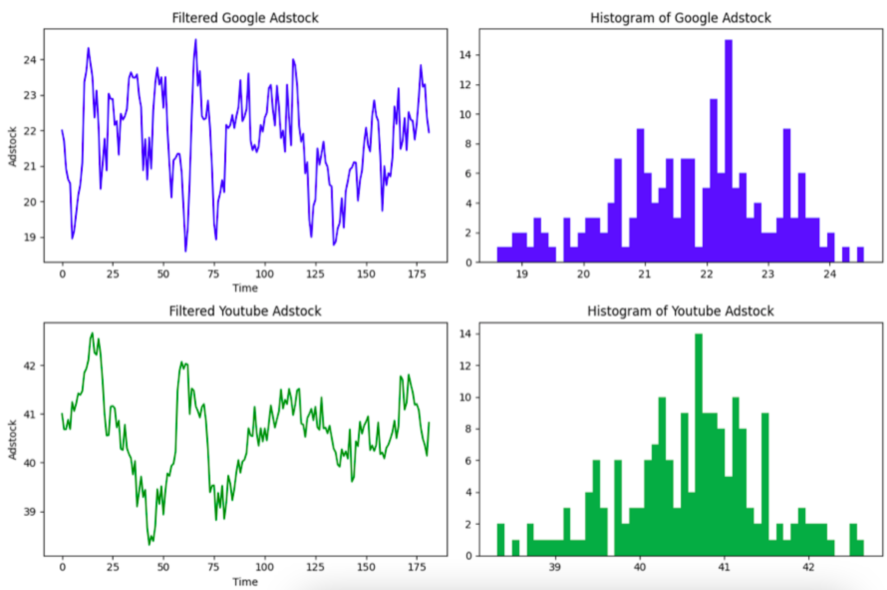

Figure 1: Plots of Filtered Google and Youtube Adstocks

The filtered Google adstock fluctuates around a certain range between 19 and 24 units. The time series shows regular variations, but the movement doesn't appear to follow a clear upward or downward trend over time. Accordingly, in the histogram plot of Google adstock can be seen that the bulk of the observations are centered between 19 and 24. Furthermore, the histogram appears slightly left-skewed with a skewness value of -0.30. In the histogram can be seen that a slight larger cluster of observations is present between

22 and 24 then on the symmetrical left side of the plot. Whereas, the left tail of the plot the appears to be slightly longer than the tail on the right side.

Similar to the Google adstock, the filtered YouTube adstock fluctuates within a range between 38 and 42 units. The time series seems a bit smoother, possibly indicating less volatility in comparison to the Google adstock. The histogram of YouTube adstock is slightly left skewed as well, with a skewness value of -0.24. This indicates the distribution of YouTube adstock to be less skewed than the distribution of Google adstock. In the histogram of YouTube adstock it is observed that the values tend to cluster slightly more right of the middel, around 41, than symmetrical left of the middel, around 40, which is in line with left skewness.

3.3

We have estimated the following model using the Ordinary Least Squares (OLS) regression method.

$$s_t = \mu + \phi_g g\hat{ad}s_t + \phi_y y\hat{ad}s_t + \epsilon_t \tag{1}$$

The estimated parameters are presented in Table 1.

Variable	**Coefficient**	**Std. Error**	**t-value**	**p-value**
Constant	-27071.25	844.719	-32.048	$<$0.001
Google Ads (ϕ_g)	420.301	13.223	31.785	$<$0.001
YouTube Ads (ϕ_y)	527.679	20.035	26.338	$<$0.001

Table 1. OLS Regression Results of Model (1)

The coefficient for the intercept is statistically significant. This negative constant suggests that if no Google or YouTube ad spending occurs, the model predicts that sales are going to be negative. In practical terms, this would imply that some base level of ad spending is crucial to achieve positive sales.

The estimated coefficient for Google Ads is highly significant. A one-unit increase in Google Ads adstock is associated with an increase in sales of 420.301 units, assuming other factors remain constant. The standard error for this estimate is 13.223, suggesting that the variation around the true value is relatively small.

The estimated coefficient for YouTube Ads is statistically significant. Specifically, a one-unit increase in YouTube Ads adstock corresponds to an increase in sales of 527.679 units, assuming other factors remain constant. The standard error for this estimate is 20.035, which shows the estimate is also reasonably precise but has slightly more variability than the Google Ads coefficient.

3.4

We will 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. Those numbers represent the equilibrium levels of adstock that the company would maintain under their fixed investment strategy.

Google adstock

First, we start with the basic equation

$$gads_t = \frac{\alpha \cdot 250}{1 + \delta \cdot gads_{t-1}} + \beta \cdot gads_{t-1}$$

where we assume that adstock converges

$$\lim_{t \to \infty} gads_t = gads_{t-1} = gads_\infty$$

Therefore, this gives us the following steady-state equation.

$$gads_\infty = \frac{\alpha \cdot 250}{1 + \delta \cdot gads_\infty} + \beta \cdot gads_\infty$$

Now we substitute given parameters into the above equation and multiply both sides by $1 + 10 \cdot gads_\infty$ to eliminate the denominator.

$$gads_\infty \cdot (1 + 10 \cdot gads_\infty) = 500 + 0.9 \cdot gads_\infty \cdot (1 + 10 \cdot gads_\infty)$$

After expansion, re-arranging terms and simplification we arrive to the following equation.

$$gads_\infty^2 + 0.1 \cdot gads_\infty - 500 = 0$$

Then the solution is found using the quadratic equation.

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

The final result is the positive number because we disregard the negative result.

$$gads_\infty \approx 22.31$$

Youtube adstock

The derivation for the limit of Youtube adstock is very similar. Therefore, we skip the explanation of the identical steps and only present the relevant equations.

We set again $yads_t = yads_{t-1} = yads_\infty$, so

$$yads_\infty = \frac{\alpha \cdot 250}{1 + \delta \cdot yads_\infty} + \beta \cdot yads_\infty$$

After elimination of denominator by multiplying by $1+10 \cdot yads_\infty$, expansion, re-arrangement of terms and simplification we arrive to the following result.

$$0.15 \cdot yads_\infty^2 + 0.03 \cdot yads_\infty - 250 = 0$$

After solving the quadratic equation, the final result is the positive number because we disregard the negative result.

$$yads_\infty \approx 40.72$$

Additionally, the f solve function from the library scipy.optimize comes handy in order to find the steady-state value through programming in Python.

3.5

Figure 2 shows a 2x2 grid of plots, where the left column contains plots of daily sales response to the effect of spending an extra 100 euro, in excess on the fixed budget of 250 euro, on Google Ads and Youtube Ads. The right column contains the corresponding cumulative sales response for both Google Ads and YouTube Ads.

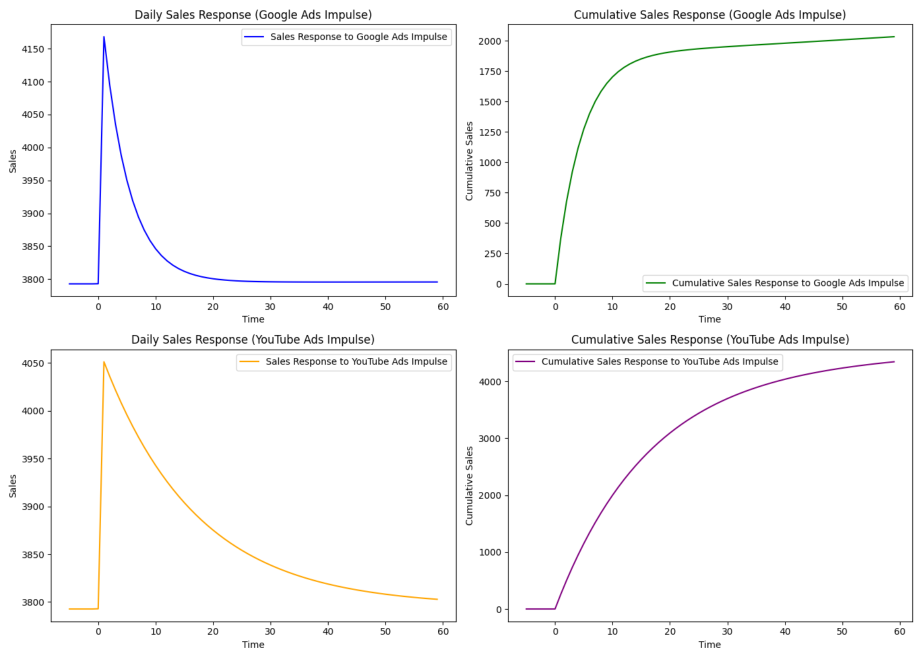

Figure 2: Plots of Daily and Cumulative Sales Response Functions

In the left column of figure 2 is observed that in the periods before the impulse the equilibrium sales is equal to 3800, corresponding to the fixed daily budget of 250 for both Google and YouTube.

Further, in the daily sales response plot for Google Ads can be seen that an increase of 100 euros in Google Ads immediately rises the daily sales in the next time point to just above 4150 euros. In the following time points the daily sales sharply declines until it reaches the equilibrium level of 3800 again at just after time point 20. Corresponding to this, in the cumulative sales response plot of Google Ads a short period of sharp increase of the cumulative sales is observed, which stagnates at around 2000 euros just after time point 20.

Subsequently, in the daily sales response plot of YouTube is observed that an impulse of an extra 100 euros spending on YouTube

Ads results in an increase to 4050 euros of daily sales in the next time period. After this period the daily sales smoothly decreases back to the equilibrium level of 3800 at around time point 60. Accordingly, in the cumulative sales response plot of YouTube can be seen that the cumulative sales for the YouTube impulse of 100 euros extra spending results in a increase smoothly to an equilibrium above 4000 euros at time period 60.

Hence, while an increase of an extra 100 euros spending on Google Ads results in an initially larger increase of daily sales compared to YouTube Ads, the effect on sales diminishes much faster back to the equilibrium of 3800 euros, where for the YouTube impulse the daily sales reduce at a much slower rate. The results of this can be seen in the differences between the final cumulative response equilibria for the Google Ads impulse response and the Youtube Ads impulse response. For Google Ads the extra 100 euro spending increases the sales cumulatively by 2000 euros, where for YouTube the final effect is much larger at an increase of above 4000 euros of extra sales. This result is in line with the coefficient estimates of question 2, where Google has a slightly lower β ($\hat{\beta}_g = 0.9$), than YouTube ($\hat{\beta}_y = 0.97$), indicating a less strongly persistent effect of ads expenditure on sales for Google.

3.6

AB testing leverages randomized experiments to generate experimental data, enabling us to uncover causal relationships between variables. It produces exogenous, *ceteris paribus* shifts in the regressor (ad spend), isolating the causal effect of a change in one variable on another while keeping all else constant. In this assignment, the shocks introduce an additional 100 euros of ad spend in each adstock equation. This randomization allows for accurate estimation of causal coefficients by isolating the impact of ad spend on sales.

AB testing plays a key role in mitigating endogeneity in advertising effectiveness studies. Without AB testing, coefficients such as ϕ_g and ϕ_y—which measure the impact of ad spend on sales—may be biased due to endogeneity, primarily from simultaneity, where factors influence both ad spend and sales. This simultaneity makes it challenging to separate the causal effect of ads on sales from the reverse effect, leading to potential misinterpretations of ad effectiveness. Without AB testing, the model's results would reflect only correlation between the dependent and independent variables, not causation.

In this assignment, AB testing is crucial for assessing ad spend effectiveness on sales. Without it, we could not confidently determine whether increased ad spending translates to higher sales, making it difficult to evaluate the investment's effectiveness.

PART II: DYNAMIC PRICING

3.7

Figure 3 shows a 2x2 grid of plots on pricing data for travel insurance of Solara company, the plots represent the data of sales, prices, acquisition costs and marketing expenditures over time. On the x-axis we see the date in years from 2005 until 2024. On the y-axis we see the value of prices, costs, marketing expenditures in euros and the value of sales in amount sold travel insurances. The blue line is corresponding to the sales, the green line is corresponding to the prices, the red line is corresponding to the Acquisition Costs and the purple line is corresponding to the marketing expenditures.

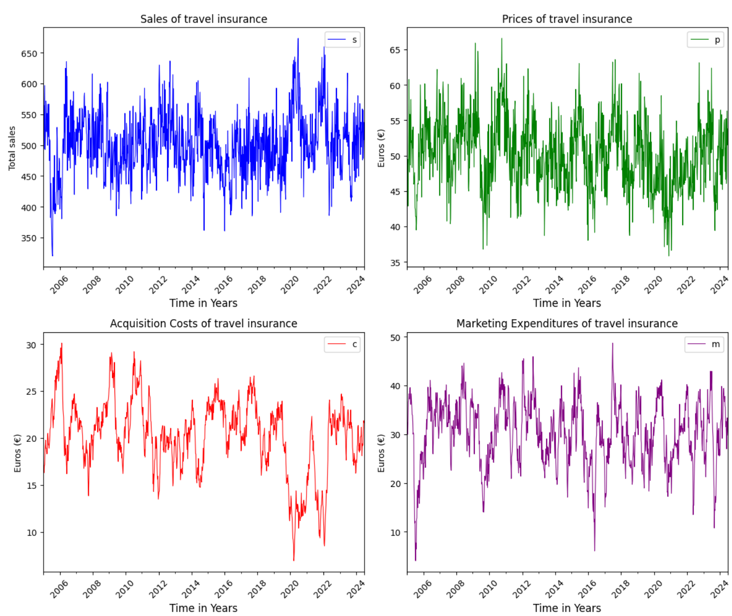

Figure 3: Plots on Pricing Data for Travel Insurance

3.8

We ran a simple OLS regression of pricing p_t on sales s_t, as can be seen in the following equation

$$s_t = \alpha + \beta p_t + \varepsilon_t. \tag{2}$$

We obtained the following parameters:

$$\hat{\alpha} = 414.5996 \quad \text{and} \quad \hat{\beta} = 1.7225.$$

Both parameters were statistically significant with a P-value < 0.05.

3.9

Predictive modeling is relevant when your sole objective is obtaining accurate predictions or forecasts. As demonstrated earlier, the estimated coefficient $\hat{\beta}_T = 1.7225 > 0$ suggests a positive relationship between s_t (sales) and p_t (price). This relationship is illustrated in Figure 4. Such a result is expected as prices and sales often co-move and exhibit positive correlation; prices react to sales. When demand is high the prices will typically increase and when the demand is low the prices will decrease. For predictive modeling we need an estimate that describes how an increase in price affects sales, it is not relevant if there is endogeneity. Therefore, this model is indeed useful as a predictive model for sales. It is however important to note that the R^2 is low (0.027) this indicates a poor fitted model.

Figure 4 displays the linear regression of sales on prices, on the x-axis we have the selling price and on the y-axis we have the total sales of travel insurances. The red line is visualizing the linear relationship, this is a positive correlation.

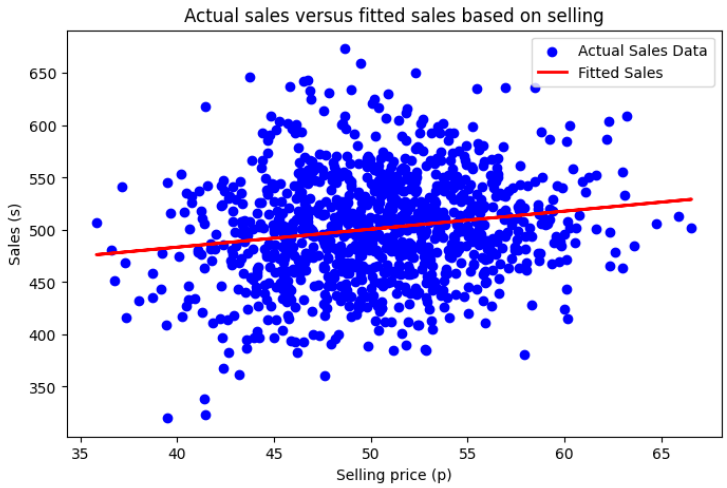

Figure 4: Plots on Pricing Data for Travel Insurance

If our estimator for $\hat{\beta}_T$ is consistent, meaning it will converge in probability to the true value β_0 as $T \to \infty$, where T represents the sample size. This β_0 will deliver the best predictive accuracy for the model, but does not correctly describe the causal impact of prices on sales. If the model is correctly specified, this limit will be the true data generating process. If the model is not correctly specified, this limit will be the pseudo-true parameter θ_0. This is for the least-squares extremum estimator equal to the closest distance in Ł$_2$ norm to the true data generating process.

3.10

Structural modeling is needed when we wish to understand the causal relation between variables. For our regression, it is the causal effect of a change in price pt on the expected sales st, keeping all the other factors constant: exogenous ceteris-paribus. It is crucial to note that this regression is not appropriate for this structural relationship and thus inadequate for dynamic pricing due to the positive estimate $\hat{\beta}_T = 1.7225 > 0$. This implies that a higher price will give a higher demand in travel insurances. This contradicts the intuitive understanding that consumers are more likely to buy more when prices are lower and vice versa. Therefore, this model is not a good structural model. This problem occurs because the prices are simultaneously determined: pt affects st, but st affects pt as well. Further is the sales equation missing additional explanatory variables who affect sales such as marketing expenditures which leads to endogeneity.

3.11

Because we are concerned that price p_t might be endogenous due to simultaneity, we decided to follow the instrumental variable approach using the following regressions:

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

and

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

The goal is to find the structural-causal relationship between sales and prices.

In the first regression, where we regressed prices p_t on costs c_t, we obtained the following estimates:

$$\hat{p}_t = \delta + \gamma \cdot c_t + u_t,$$

with

$$\hat{\delta} = 38.8635 \quad \text{and} \quad \hat{\gamma} = 0.5646.$$

Here, $\hat{\delta}$ represents the intercept (constant), and $\hat{\gamma}$ is the estimate of the effect of costs on price. Both estimates are significant.

In the second regression, where we regressed sales s_t on the predicted price $\hat{p}_t$, the estimates are as follows:

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

with

$$\hat{\alpha} = 1005.3320 \quad \text{and} \quad \hat{\beta} = -10.0232.$$

In this regression, $\hat{\alpha}$ represents the intercept, and $\hat{\beta}$ is the estimate of the effect of the predicted price on sales. Both estimates are significant with a P-value of 0.000.

To measure the causal impact of an increase in price p_t on expected sales, the instrumental variable condition must hold:

1. **Relevance**: The instrument (costs, c_t) must be correlated with the endogenous variable (prices, p_t). This was tested in the first-stage regression, where the coefficient $\hat{\gamma}$ was found to be significant and the F-statistic had a value of 263.3 (>10) indicating that costs are indeed a relevant instrument.

2. **Exogeneity**: The instrument must not be correlated with the error terms in both the first and second stages. Specifically, the instrument must be uncorrelated with both u_t and ε_t. This is expressed in the following condition:

$\epsilon_t \perp c_t \quad \text{and} \quad u_t \perp c_t$

Meaning the instrument must be exogenous to both error terms. From the exogeneity tests in python it holds that ct is exogenous, suggesting that the costs c_t are a valid instrument.

Given that both the relevance and exogeneity conditions hold, we can now calculate the causal impact of an increase in price p_t on expected sales. According to the second regression, a one-unit increase in price, while keeping all the other factors constant, leads to a decrease in sales by -10.0232, which is the value of $\hat{\beta}$.

3.12

With the Hausman-Durbin-Wu test we can test for endogeneity. Let $\tilde{\theta}_T$ (regression with instrumental variable) and $\hat{\theta}_T$ (regression without the instrumental variable) be asymptotically normal to the true parameter θ_0.

$$\sqrt{T}(\tilde{\theta}_T - \theta_0) \xrightarrow{d} N(0, \tilde{\Sigma}) \quad \text{and} \quad \sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, \hat{\Sigma}),$$

as $T \to \infty$. Furthermore, let $\tilde{\theta}_T$ and $\hat{\theta}_T$ satisfy

$$\tilde{\theta}_T \xrightarrow{p} \theta_0 \quad \text{and} \quad \hat{\theta}_T \xrightarrow{p} \theta_0,$$

as $T \to \infty$.

Under the null hypothesis we have that both regressors are exogenous and consistent. Under the alternative hypothesis we state that only one regressor is consistent under endogeneity: the regression with an instrumental variable $\tilde{\theta}_T$. The test is based on the difference between these two estimators:

H0:

$$\sqrt{T}(\tilde{\theta}_T - \hat{\theta}_T) \xrightarrow{d} N(0, \tilde{\Sigma} - \hat{\Sigma}),$$

HA:

$$\sqrt{T}(\tilde{\theta}_T - \hat{\theta}_T) \xrightarrow{p} \infty,$$

Note here that the scaled asymptotic difference under the alternative hypothesis will go to infinity as T goes to infinity. As a result, we can reject the null hypothesis of regressor exogeneity for large values of the test statistic.

Our test value for the Hausman-Durbin-Wu:

$$422930.014$$

with a P value of 0.000. Therefore, we reject our null hypothesis that both regressors are exogenous and consistent. And we accept our alternative hypothesis that only the regressor with the instrumental variable is consistent under endogeneity.

3.13

For question 13 we added marketing expenditures to the model to obtain the following second stage of the 2SLS:

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

Where we also included the market expenditure in the first stage LS to determine $\hat{p}_t$ to prevent biased estimation of $\hat{p}_t$.

$$\hat{p}_t = \delta + \gamma c_t + \chi m_t + u_t$$

The regression outcomes are listed below

Table 2. Regression Estimates

Variable	**Coefficient**	**Standard Error**	**P-value**
Regression of Prices on Costs Only ($p_t = \delta + \gamma c_t + u_t$)			
Constant (δ)	38.8635	0.732	0.000
Cost (γ)	0.5646	0.035	0.000
Regression ($s_t = \alpha + \beta \hat{p}_t + \psi m_t + \varepsilon_t$)			
Constant (α)	879.1832	25.567	0.000
Predicted Price (β)	-10.2632	0.500	0.000
Marketing (ψ)	4.5729	0.166	0.000

From the regression results we can see that marketing has a positive and significant effect on the sales. We could have chosen to include marketing as a control variable also into the first stage, however we chose not to do this, as it was stated in the text that pricing and marketing expenditure were evaluated separately at random.

3.14

For dynamic pricing, we need an estimate that describes how an exogenous ceteris-paribus increase in pt affects st.

We would like to optimize/maximize the profit als a function of p_{T+1}:

$$\pi_{T+1} = s_{T+1}(p_{T+1} - c_{T+1}) - m_{T+1}$$

using the function described in question 13, we can describe sales as

$$s_{T+1} = \alpha + \beta p_{T+1} + \psi m_{T+1}$$

where $p_{T+1} = \hat{p}_{T+1}$

We see that the profit also depends on the sales and the cost, which are to be determined for 1 step ahead using simple AR(1) models;

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

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

This renders the following expectations;

$$E(c_{T+1} \mid F_T) = \mu_c + \phi_c c_T$$

$$E(m_{T+1} \mid F_T) = \mu_m + \phi_m m_T$$

Substituting these expectations into the profit function $E(\pi_{T+1}|F_T)$ gives the following expression;

$$\pi_{T+1} = (\alpha + \beta p_{T+1} + \psi m_{T+1}) \cdot (p_{T+1} - (\mu_c + \phi_c c_T)) - (\mu_m + \phi_m m_T)$$

To optimize this expression as a function of pricing, we can take the derivative of the function with respect to pricing and set it to zero, to obtain

$$\frac{\partial \pi_{T+1}}{\partial p_{T+1}} = \alpha + 2\beta p_{T+1} + \psi m_{T+1} - \beta\left(\mu_c + \phi_c c_T\right)$$

Solving this equation for p_{T+1} gives the optimal price:

$$p_{T+1}^{*} = \frac{\beta\left(\mu_c + \phi_c c_T\right) - \alpha - \psi m_{T+1}}{2\beta}$$

The results of the AR(1) regressions are described below graphicx

Table 3. AR(1) Processes: Costs and Marketing Expenditures

Variable	Costs			Marketing		
	Coefficient	**Standard Error**	**P-value**	**Coefficient**	**Standard Error**	**P-value**
Constant (μ_c, μ_m)	0.6819	0.167	0.000	3.2398	0.438	0.000
Lagged (ϕ_c, ϕ_m)	0.9664	0.008	0.000	0.8927	0.014	0.000
R-squared		0.934			0.797	
Adjusted R-squared		0.934			0.797	
F-statistic		1.432e+04			3987	
Prob (F-statistic)		0.00			0.00	

In Table 3 the expected values of the cost and market expenditure according to the AR models are described. The last two rows indicate the optimal price evaluated at T+1. We can observe that the price can be increased significantly to obtain optimal profit. The last observed price was 51,54 and the optimal price at T+1 is 60,27. This is a difference of 8,73.

Table 4. Expectations and Optimal Price

Variable	**Expected Value**
$E(c_{T+1})$	21.5843
$E(m_{T+1})$	29.8321
Optimal Price at $T+1$	60.2705
Last Observed Price	51.5381