FINAL PROJECT

HU WANG

Contents

1. Introduction

1.1. **Overview.** The goal of this article is to study and optimize an investment for online advertising according to age group distribution. We also study the effects of a risk penalty factor on our optimal solution.

1.2. **Motivation.** Our motivation is that online advertisement is a growing industry with high revenue, but investments always comes with risk, and we would like to study the optimal investing plan.

1.3. **Specific Questions in this Line of Research.** We'd like to study the distribution of four social media users by five age groups, and looking for the best optimal solution for investors. We also want to avoid investors focus on some specific age groups, which might lead to creating negative effects and neglect other age groups, Therefore we want to find a risk penalty factor that can get us a ideal optimal solution.

2. Models

2.1. **Basic Model.** Our basic model is an linear program maximizing the revenue, and finding the best way to divide our principal. We maximize

$$\sum_{i=1}^{4}\sum_{j=1}^{5} u_{ij}x_{ij} \tag{1}$$

where

$$x_{ij} = [0,1] \quad \text{the proportion of our investment in media i, and age group j} \tag{2}$$

$$\sum_{i=1}^{4}\sum_{j=1}^{5} x_{ij} = 1 \tag{3}$$

i.e. Our total investment sums up to 1.

To reach a wider audience for our advertisement, we need to constraint the money we spend on each age group. In our case, we would like to bound the money spent on our two highest utility age group for all social media sites by at least 50 percent, and at most 70 percent of our principal; since we want to invest more on higher utility age groups but not too much; and at the same time make sure we are also investing at least 1 percent on other age groups. Hence we introduce two new constraint.

$$0.5 \leq \sum_{i=1}^{4}\sum_{j=1}^{2} x_{ij} \leq 0.7 \tag{4}$$

i.e. Our investment in the first two age groups must be more than 50 percent, less than 70 percent.

$$x_{ij} \geq 0.01 \tag{5}$$

i.e. We want all x_{ij} to have at least 1 pervent of our investment.

This will be our basic linear program.

2.2. **More Refined Model.** We would like to introduce a parametric risk penalty factor so that the linear program penalizes us if our proportion of spending is too dominant on the younger age groups. We want to reach a wider range of audiences by dividing a fair amount of our principal towards the elder age groups. Our risk penalty factor is based on the statistical variance of the data we have. This will allow us to minimize the difference in proportion we spent on each age group, hence penalizing our optimal solution if it is dominated by certain age groups. We start by introducing four new variables:

$$y_j = \sum_{i=1}^{4} x_{ij} \tag{6}$$

i.e. which is the total amount of invested proportion in age group j

$$z_i = \sum_{j=1}^{5} x_{ij} \tag{7}$$

i.e. The total amount of invested proportion in social media i.

Then we have our α_j and β_i which is a scalar factor for the penalty factor of age groups and social medias respectively. α_i and β_j will have different weights, and we will set the weights to $\alpha_5 \geq \alpha_3 \geq \ldots \geq \alpha_1$ and $\beta_4 \geq \beta_3 \geq \cdots \geq \beta_1$. We want the weight of the younger age groups to be less because the proportion of money invested will be greater, while the weight of the elder age groups will be higher since the money invested will be lesser. This will then help us balance out the proportion of each age group. Similar with α, β is weighed accordingly.

Hence our new utility function will be

$$u^* = \sum_{i=1}^{4}\sum_{j=1}^{5} u_{ij}x_{ij} - [\sum_{j=1}^{5}(y_j - \frac{1}{5})^2\alpha_j] - [\sum_{i=1}^{4}(z_i - \frac{1}{4})^2\beta_i] \tag{8}$$

keeping the constraints (3), (4), (5).

This will give us a optimal solution that has a balanced spending among each age group and social media.

2.3. **Modeling Our risk penalty factor.** A refinement of the above model will be to find the optimal α and β by scaling them from 0 to 100 and testing the optimal solution of our utility function. This model will keep track of how α and

β affects the optimal solution.

We first set β to 0 and test the utility function

$$u^* = \sum_{i=1}^{4}\sum_{j=1}^{5} u_{ij}x_{ij} - [\sum_{j=1}^{5}(y_j - \frac{1}{5})^2\alpha_j] \tag{9}$$

i.e. we scale α from 0 to 100.

This will allow us to find an optimal number for α such that the proportion of our principal is balanced. Similarly we can test β the same way with a new utility function,

$$u^* = \sum_{i=1}^{4}\sum_{j=1}^{5} u_{ij}x_{ij} - [\sum_{i=1}^{4}(z_i - \frac{1}{4})^2\beta_i] \tag{10}$$

i.e. we scale β from 0 to 100.

3. Data

We have the age distribution of social media from Statista.com, users as of 3rd quarter of 2014. But we don't want the utility to be a percentage, we multiply the distribution by 100 as our utility. [1]

4. Our Algorithms

We implemented our project in Python with Gurobi libraries gurobipy to formulate and solve models as described in previous sections. Starting with the basic model, we created a Model variable in and initialize the decision variables with addVars. At the same time set the lower bound to be 0.01 to satisfy our constraint. Then we constructed the objective function by using setObjective. Then we used addConstr to set the constraints described in our basic model and visualize with

[1] Refer to Appendix A

printSolution(). For advanced models, we have two quadratic risk variables to calculate and store the two quadratic terms and then reset the objective function by subtracting these two terms. And re-optimize the program.

5. Our Main Current Results

5.1. **Results from Testing our Model.** We started by testing the affect of each α and β as we shift it from 0 to 100, and we found out that the effects of α and β are identical[2]. After we tested the trends of α and β we test and find an optimal solution for the spread of our investment. We set α_1 to α_5 and β_1 to β_5 to different and increasing numbers, and got a satisfied trend of spread throughout when $\alpha_1 = 10$, $\alpha_2 = 30$, $\alpha_3 = 50$, $\alpha_4 = 70$, $\alpha_5 = 90$, $\beta_1 = 10$, $\beta_2 = 20$, $\beta_3 = 20$, and $\beta_4 = 30$.[3] We tested it again without the constraints from the basic model, and got a better optimal solution, where $\alpha_1 = 0$, $\alpha_2 = 20$, $\alpha_3 = 40$, $\alpha_4 = 60$, $\alpha_5 = 80$, $\beta_1 = 0$, $\beta_2 = 10$, $\beta_3 = 10$, and $\beta_4 = 20$. Then we tried the values on its own to see if only α or β is sufficient to spread out our investment. As expected the results were bad, and we concluded that α and β must both be in use in order to get a optimal spread in our investment.

5.2. **Conclusion of the Results.** The first conclusion we have reached is the importance of including both α and β in the equation to make a spread such that we can target the audience we want. From our final optimal solution, we can see that our focus on the younger age groups (16-34) are not neglected, with about 57 percent of our total investment. At the same time, we have a 20 percent investment

[2]Refer to Appendix B

[3]Refer to Appendix C

in the elderly age groups (45-64) with 8 percent in social media Facebook. Without the inclusion of both α and β the spread would not even come close to what we want.

The values we set for each α_j and β_i is also a crucial part of our results, and after several tests we can conclude that the difference between these nine values is what dominates how the optimal solution spreads. This difference in the weights of α and β allows us to conclude that the punishment on younger age groups must be smaller, to protect our main investment in the younger age groups. Thus the values of αs and βs form an increasing trend. Overall we found out that these values successfully spreads our optimal solution without the use of other constraints.

6. Obstacles Encountered So Far

(1) One of the obstacles we encountered is deciding the difference between each α_j and β_i. We did multiple tests to see how much each should weight, and how effective would it be.

7. Questions for Future Research

(1) From the data above, as α and β increases, the change in optimal solution slows down. For further research, if we set the α and β to infinity, we might get an interesting limit of the optimal distribution for each small group, and this limit can be studied more.

(2) The superposition principle of two risk penalty factors. The superposition principle is linear in our model. However, if we change it to a nonlinear superposition, how will the new model be? The new model can be a multiplication of risk penalty factor.

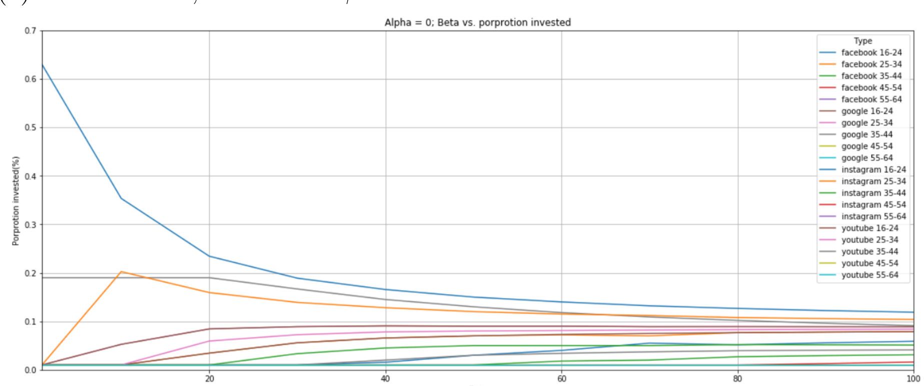
8. Appendices

8.1. **Appendix A.** Utility chart.

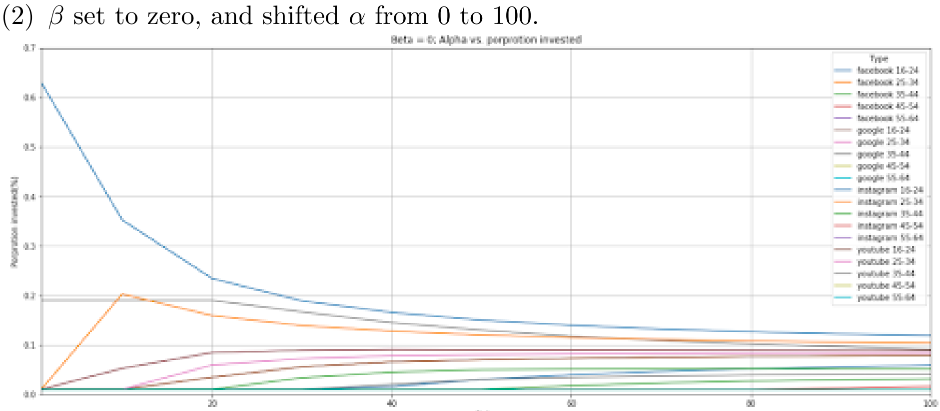
	16 - 24	25 - 34	35 - 44	45 - 54	55 - 64
Google	29	31	20	13	7
Youtube	31	30	20	13	7
Facebook	25	29	22	15	9
Instagram	37	34	18	8	3

8.2. **Appendix B.**

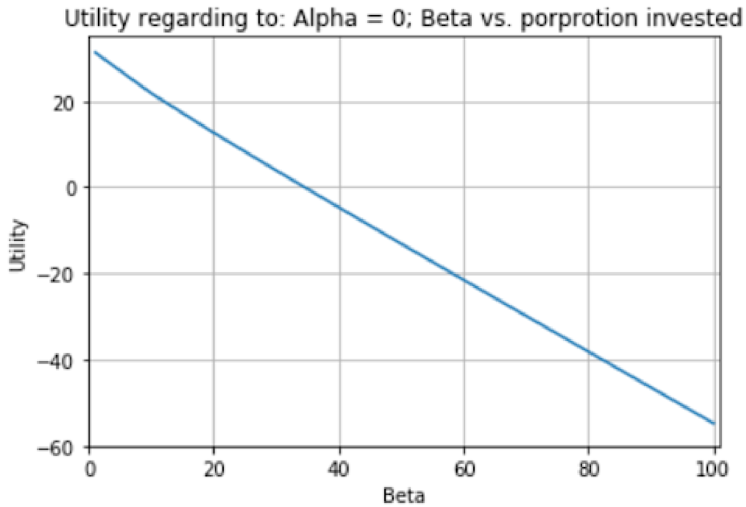
(1) α set to zero, and shifted β from 0 to 100.



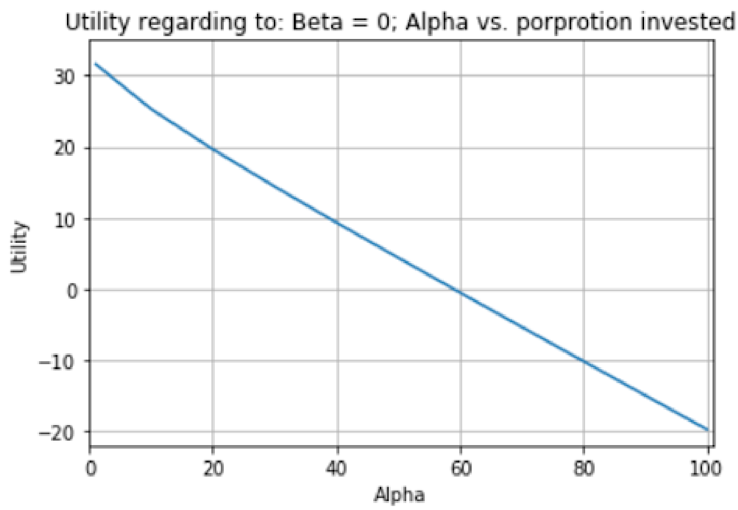
(2) β set to zero, and shifted α from 0 to 100.



(3) Utility trend for α set to zero, and shifted β from 0 to 100.

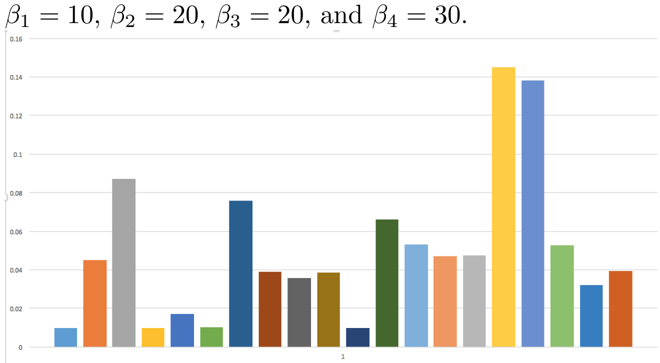


(4) Utility trend for β set to zero, and shifted α from 0 to 100.

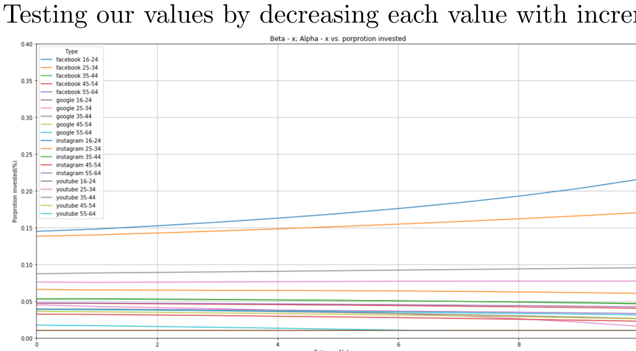


8.3. Appendix C.

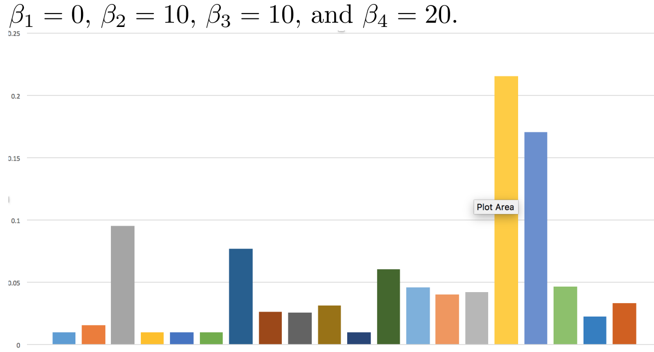
(1) Spread of investment when $\alpha_1 = 10$, $\alpha_2 = 30$, $\alpha_3 = 50$, $\alpha_4 = 70$, $\alpha_5 = 90$, $\beta_1 = 10$, $\beta_2 = 20$, $\beta_3 = 20$, and $\beta_4 = 30$.



(2) Testing our values by decreasing each value with increments of 1.



(3) Spread pf investment when $\alpha_1 = 0$, $\alpha_2 = 20$, $\alpha_3 = 40$, $\alpha_4 = 60$, $\alpha_5 = 80$, $\beta_1 = 0$, $\beta_2 = 10$, $\beta_3 = 10$, and $\beta_4 = 20$.



The two most invested bars(yellow and blue) are for the age groups 16-24 and 25-34 in social media Instagram.