Industrial Organization and Data Science

Instructors: Justin Rao, Affiliate Professor & Senior Researcher, Microsoft

Jacob LaRiviere, Affiliate Professor & Senior Researcher, Microsoft

Emails: justin.rao@microsoft.com

jlariv@microsoft.com

Course Assignments & Reading

Course assignments should be printed (code, output and descriptive answers) and turned in at the start of class unless otherwise noted. Feel free to work in groups but everyone is required to turn in their own work with answers written in your own words. In both calculations and complex ideas, write down each step of logic used in reaching your conclusion. Keep in mind that in most cases a good answer is one precise sentence; quality is heavily favored over quantity. This will be graded on a full credit, half credit and no credit basis. All work must be typed

Discussion questions do not need be written out ahead of time. At the beginning of each class the professors will lead a discussion around these questions. Students will be called on, potentially at random, to add their insight. This part of class will contribute heavily to your course participation grade.

Week 8, due May 25

Assignment to be turned in. Please turn in your R output and answers to the questions.

In this assignment we're going to do two things. First, we're going to take the output from the week 7 assignment, interpret the coefficients and turn it into an excel spreadsheet to create a pricing and revenue "model". Second, we're going to try to put the data into a panel format and first difference the prices.

1) The output of the week 7 homework was a series of 3x3 elasticity matrices. If you had trouble with your own code, you can run the code I sent out. For example, the elasticity matrix should look like this:

	Dom	MinMaid	Trop	leaf
Dom2	-2.999234	0.9987842	-0.438856	2
MinMaid2	0.5892045	-2.886916	0.5446749	2
Trop2	0.2406747	0.2673355	-2.794327	2
Dom4	-2.817939	1.1613242	-0.19022	4
MinMaid4	0.6421868	-2.357694	0.6188461	4
Trop4	0.1932227	0.3839389	-2.106839	4
Dom5	-3.121804	0.8630777	-0.347788	5
MinMaid5	0.2132359	-2.028908	0.3154762	5
Trop5	-0.013042	0.1717764	-1.979717	5

   a. Looking at the code sent out for week 7 HW, interpret that coefficient of in location [5,1] of the matrix (e.g., .642). Interpret the coefficient in location [6,1]. What can you conclude about Dominick's in Leaf 4?

i. Do the same for leaf 5 (e.g., [8,1] versus [9,1]).
   1. What can you conclude about stores in higher income areas?
   2. How should this impact pricing in leaf 4 versus leaf stores.

b. Consider the following regression in line 88 of the code:

```
regm =
lm(logmove.minute.maid~log(price.dominicks)*feat.dominicks+log(price.m
inute.maid)*feat.minute.maid+log(price.tropicana)*feat.tropicana,
data=temp_df)
```

i. Interpret each coefficient in this regression. Your answer should ten one line answers.
ii. Run this regression for leaf four. What coefficients are not significant?

2) Let's create a model with our coefficient estimates for leaf 4 in excel. You'll also need to pull out the average quantities for MM, Dom and Trop sales by store for a baseline. For this exercise we will completely ignore featuring brands as an option.
   a. For costs, assume that the minimum observed price in the dataset for Dom, MM and Trop is the cost (don't restrict to only leaf four for this part; use only leaf four for the rest).
      i. This is purely a modelling decision and isn't coming from anywhere.
   b. Using that cost, baseline sales and average prices, what is the total profit on average for each brand?
   c. Using your elasticity estimates, what would the change in profits in the following 6 scenarios (e.g., each of these experiments independently). Make sure to account for both own price and cross price elasticities:
      i. Increase price of Dom by 10%.
      ii. Decrease price of Dom by 10%.
      iii. Increase price of MM by 10%.
      iv. Decrease price of MM by 10%.
      v. Increase price of Trop by 10%.
      vi. Decrease price of Trop by 10%.
   d. Do any of these increase profits? If so, can you come up with anything else which would even better?
   e. Now assume that an orange fungus increases the price of all orange juice by 25%.
      i. Use your excel sheet pricing model to make a suggestion.
3) Now consider trying to put your data in a "panel data" format. The goal is to move toward estimating intertemporal elasticities.
   a. Interpret the coefficients of this proposed regression, but do not estimate it, assume *I* is a store, *j* is a product (e.g., minute maid) and *t* is a week:

$$\ln(Q_{ijt}) = \alpha + \ln(Q_{ijt-1})\beta_1 + \ln(Q_{ijt-2})\beta_2 + [\ln(P_{ijt}) - \ln(P_{ijt-1})]\gamma_1 + [\ln(P_{ijt-1}) - \ln(P_{ijt-2})]\gamma_2 + \epsilon_{ijt}$$

   b. Restrict the same to only minute maid quantities and prices.
   c. Now try to put the data in a panel format. There are a couple of ways to do this and here is one.
      i. Create a new variable called `df$week.lag=df$week-1`.
      ii. Save a new dataframe called `df_temp <- df`
      iii. You can then the two dataframe (`df_temp` and `df`) into a single dataframe by store-week and store-week.lag

iv. Do this one more time and run the above regression. Is there evidence of intertemporal substitution?

BONUS: Add in other variables which could be used then run a LASSO using the `glmnet()` package.