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 7, due May 18

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

In this assignment we're going to "level set" (corporate slang!) by determining which demographic characteristics explain price differences across stores, then use that information to estimate different demand curves by store-level demographic characteristics. Effectively, this puts stores into bins (e.g., market "types") for which we'll estimate separate demand curves and set pricing policy.

1. Create a sales weighted price for orange juice by store.
    a. You can use the weighted.mean() function for each store-week combination
    `weighted.mean(oj$price, oj$percent_sold)`
    Before doing so, though, you'll need to create the percent of sales by each brand within a store. On way is to use the `group_by` function in the `dplyr` package.
    An alternative is to use the `aggregate` function can help you with that after you take `mydata$Q <- exp(mydata$logmove)` to get sales in levels.

    ```
    temp <- aggregate(oj[,column_location], list(oj$store,
    oj$week), sum)
    ```

    This last command will create store OJ sales sums from which you can calculate `oj$percent_sold` after merging it back in to the original dataset.

2. Now use `oj$weighted_price` as the LHS variable in a regression tree to predict differences in sales weight prices with store demographics as RHS variables.

   a. Play around with a couple different complexity parameters to get a feel for the data
   b. Choose three different leaves to group stores into based upon what explains sales weighted price.
      i. Assign each store to one of these leaves (we used this code previously).
3. Estimate the own price elasticities for each one of the store buckets/leaves using the preferred specification:

   ```
   reg_int <- glm(logmove~log(price)*brand*feat,
   data=oj_leaf_L)
   ```

   a. Now estimate cross price elasticities jointly with own price elasticities. This means you must create a dataframe which has the prices of all types of OJ at the store.
   b. You'll also have to run 3 separate regressions for each leaf for a total of nine regressions.

      ```
      reg_int <- glm(logmove_D~log(price_D)*feat*brand +
      log(price_T)*feat*brand + log(price_MM)*feat*brand,
      data=oj_leaf_L_D)
      ```

      In this example, we are investigating the own and cross price elasticities for Dominick's brand (D) within leaf L.
      i. Save the coefficients for each leaf in a 3x3 matrix. The diagonals will be own price elasticities and the off diagonals will be cross price elasticities.
      ii. There will be a unique 3x3 matrix for each leaf.
      iii. The 3x3 matrices WON'T be upper triangular because we're estimating three unique regressions for each leaf.
   c. Comment on any differences between own and cross price elasticities by leaf.
4. Now let's use the elasticities to think about pricing differentials.
   a. In the leaf with the highest own-price elasticities, what should the markups be relative to the other leafs?
   b. How do cross-price elasticities vary with the highest versus lowest own price elasticity leafs?
      i. What does this imply about differences in markups within high versus low elasticity stores across brands?
      ii. Can you say anything about what this means for the timing of sales? Should they occur at the same or different times across stores?