# Problem Set 5

Zilin Xiong

October 19, 2021

## 1 Introduction of models and data

In this problem, we will use MSE(Maximum Score Estimator) to predict the parameters of the given functions. Packages we use are Pandas, Numpy, Scipy.optimize, and Geopy.distance. Aside from the default packages, we installed geopy by pip.install.geopy.

To start with, we look at the data set and try to clean it. I divide year 2007 and 2008 data apart, and also divide buyers and target data apart from each other. Because we need to calculate the distance between buyers and targets, I pick the longitude and latitude value from the data set. And get the distance data in the format as "buyer-lat2007,buyer-lat2007" for both buyers and targets in 2007 and 2008. Then we calculate and store the distance value.

## 2 MSE function

### 2.1 Formula and function

We define the MSE function here as below:

$$f_m(b,t) = x_{1bm}y_{1tm} + \alpha x_{2bm}y_{1tm} + \beta distance_{btm} + \varepsilon_{btm},$$

Figure 1: Formula 1

Where x1bm is the number of stations owned by the parent company of the buyer y1tm is the population in range of the target in market m x2bm is an indicator for corporate ownership, and distance btm is the distance (in miles) between the buyer and target.

We represent the observed matches as "i", and the counterfactual matches as "j". Under the direction of the function as below, we construct our function in a "for" loop.

$$\hat{\beta} = \arg\max Q(\beta) = \sum_{y=1}^{Y} \sum_{b=1}^{M_y-1} \sum_{b'=b+1}^{M_y} \mathbb{1}\left[f(b,t|\beta) + f(b',t'|\beta) \geq f(b',t|\beta) + f(b,t'|\beta)\right]$$

Figure 2: Function 1

After that, we define the boundaries of the parameters. And for different variables. Finally, we get our results from the function by using the method "best1bin". One thing to notice here is that, for year 2008 we start at 45, because 2007 has 45 values, and the 2008 data start from line 45.(line 0 is the first line)

We have the function as:

estimate-fomula[i,j] = buyer-nsb[i + start-index] * population-per-mil[j + start-index] + param[0] * buyer-cob[i + start-index] * population-per-mil[j + start-index] + param[1] * distance[i, j]

And the MSE as:

estimate-fomula[i,i] + estimate-fomula[j,j]) ¿ (estimate-fomula[i,j] + estimate-fomula[j,i]

### 2.2 Results:

For 2007: alpha, beta =

[402.69862708 771.65597602]

For 2008: alpha, beta =

[140.83911066 637.65178431]

## 3 MSE with price information

Similar to the previous section, we use the formula and function as below.

$$f_m(b,t) = \delta x_{1bm} y_{1tm} + \alpha x_{2bm} y_{1tm} + \gamma HHI_{tm} + \beta distance_{btm} + \varepsilon_{btm},$$

Figure 3: Formula 2

We have the function as:

estimate-fomula[i,j] = param[0] * buyer-nsb[i + start-index] * population-per-mil[j + start-index] + param[1] * buyer-cob[i + start-index] * population-per-mil[j + start-index] + param[2] * hhi-target[j + start-index] + param[3] * distance[i, j]

$$Q(\beta) = \sum_{y=1}^{Y} \sum_{b=1}^{M_y - 1} \sum_{b'=b+1}^{M_y} \mathbb{1}\left[ f(b,t|\beta) - f(b,t^{'}|\beta) \geq p_{bt} - p_{b't'} \wedge f(b^{'},t^{'}|\beta) - f(b^{'},t|\beta) \geq p_{b't'} - p_{bt} \right]$$

Figure 4: Function 2

This time we have four parameters, and we define them as param[0]:delta, param[1]:alpha, param[2]:gamma, param[3]:beta. And we introduce the price as price-per-mil.

And the MSE as:

estimate-fomula[i,i] - estimate-fomula[i,j]) ¿ (price-per-mil[i + start-index] - price-per-mil[j + start-index]) and (estimate-fomula[j,j] - estimate-fomula[j,i]) ¿ price-per-mil[j + start-index] - price-per-mil[i + start-index]

Finally, we get the results.

### 3.1 Results

For 2007: delta, alpha, gamma, beta =

[ 6.06814571e+02 -4.43770279e+02 -1.33165436e-01 -6.01644447e+01]

For 2008: delta, alpha, gamma, beta =

[-892.65055141 625.73495182 46.55977302 -253.11990843]