# Problem Set #5

ECON 833, Prof. Jason DeBacker
Due Tuesday, October 19, 10:05 a.m.

This problem set is designed to provide you with hands on experience implementing a maximum score estimator. As you do this you will develop your skills in Python, with particular emphasis on working with conditionals and loops.

Consider a model of a one-to-one matching market representing radio station mergers. Each year there is a national market where radio station owners target new stations. These markets are independent across years. Thus, the model will be analogous to Akkus, Cookson, and Hortaçsu (*Management Science*, 2016).

You will estimate the parameters of the payoff functions that represent the relative importance of corporate ownership and geographic proximity compared to size sorting. In particular, the payoff to the merger between radio station buyer $b$ and target $t$ in market $m$ is given by:

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

where $x_{1bm}$ is the number of stations owned by the parent company of the buyer and $y_{1tm}$ is the population in range of the target in market $m$, $x_{2bm}$ is an indicator for corporate ownership, and $distance_{btm}$ is the distance (in miles) between the buyer and target. The match-specific error term, $\varepsilon_{btm}$ is independent across matches.

Next, estimate the version of this model with transfers (the prices pay to aquire the target station). Here, you'll use the data on the of the merger and a different inequality in your score function (see your notes on Akkus, Cookson, and Hortaçsu (*Management Science*, 2016)). In this case, let the payoff function also include target characteristics (which can now be identified):

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

where $HHI_{tm}$ is the Hindahl-Hirschman Index measuring market concentration (a higher index means a more concentrated market) in the location of the target in market $m$.

### Estimation Exercise

Please estimate the parameters of the models, $(\alpha, \beta)$, and $(\delta, \alpha, \gamma, \beta)$ using the data described below and a maximum score estimator. You do not need to construct confidence intervals - just the point estimates are sufficient.

## Data description

The data are in the CSV file `radio_merger_data.csv`. There are 2 independent markets representing mergers in 2007 and 2008. There are 45 mergers in 2007 and 54 in 2008. A unique id is assigned to each buyer and target. The variable names should be mostly obvious. Prices are in constant dollars, population in number of people (you might put these into millions of dollars and people - or into logs of thousands of dollars/people - to help with scaling). You will want to restructure these data into your $f(b,t)$ arrays for estimation.

In addition, I do not give you distance between all the possible pairs of buyers and targets. Rather you are given the coordinates (latitude and longitude) or the buyer's location and the target's location. You'll need to compute the distance between these points for all combinations of buyers and targets. To do this, I suggest you use the `geopy.distance.geodesic` function in the GeoPy package.

## Hints

- The maximum score objective function is not smooth and differentiable. Therefore, you need a optimization algorithm for nonlinear and non-differentiable objective functions for estimation. A non-gradient based method, like Nelder-Meade, may work. But many who use maximum score estimators

use a method called differential evolution (Storn and Price (1997)). This can be implemented with `scipy.optimize.differential_evolution`.

- Estimation requires comparisons of "observed" matches with "counterfactual" matches. Therefore, the data array you need for estimation is different from the provided raw datasets. In the final dataset that you need to put into the objective function, there are 2,421 ($(45 \times 44 + 54 \times 53)/2 = 2421$) unique comparisons of observed and counterfactual matches. For details on the construction of a data array required for estimation, please refer to pages 8-9 of Santiago and Fox (2009). "A Toolkit for Matching Maximum Score Estimation and Point and Set Identified Sub-sampling Inference", which contains details of Mathematica implementation of matching estimator.

## Deliverable

This assignment is your most involved thus far. As such, I would like you to write your code in a text editor and run it through the terminal or iPython. You will turn in your *.py files and a pdf document that you compile in TeX where you describe the model and results (including providing some interpretation of the results). You will put all materials for the problem set in the path `/CompEcon_Fall2021/ProblemSets/ProblemSet4` on your ProblemSets branch of your fork of the class repository. You will be graded based on your results, the efficiency of your code, and the documentation and clarity you provide in your scripts (i.e., your docstrings and inline commenting, following PEP8 rules (generally), using clear variable names).