

Microeconometrics: Problem Set 1

Due date: June 8, 2017

Your answers should be produced in L^AT_EX, and should include all relevant graph and code. Code should be in the appropriate verbatim environment and properly documented. You are allowed to work in groups but you must turn in your own writeup. Submit your assignment via email to cconlon@columbia.edu.

Part 1: Guerre Perrigne and Vuong

- Generate 1000 valuations $x \sim U[0, 1]$. Recall (as derived in lecture notes) the equilibrium bid function in this case is

$$b(x) = \frac{N-1}{N}x$$

- For 500 of the valuations, split them into 125 4-bidder auctions. For each of these valuations, calculate the corresponding equilibrium bid.
- For the other 500 valuations, split them into 100 5-bidder auctions. For each of these valuations, calculate the corresponding equilibrium bid.
- For each b_i compute the estimated valuation \tilde{x}_i using the GPV result where N_i is the corresponding number of bidders in auction i

$$\frac{1}{g(b_i)} = (N_i - 1) \frac{x_i - b_i}{G(b_i)} \rightarrow x_i = b_i + \frac{G(b_i)}{(N_i - 1)g(b_i)}$$

You will want to use **ksensity** or **kdensity**. You can construct G by integrating the density (better) or using the ECDF **ecdf** or **cumul**.

- When you compute G and g use both the Epanechnikov kernel $K(u) = \frac{3}{4}(1 - u^2)\mathbf{1}(|u| < 1)$ and the uniform/rectangular kernel $K(u) = \mathbf{1}(|u| < 1)$ where $u = \frac{x_i - x}{h}$.
- For each kernel vary the bandwidths $h = \{0.5, 0.1, 0.05, 0.01\}$
- Plot x vs \tilde{x} Comment on the performance of the procedure for different choices of kernel and bandwidth.
- If you can – report which choices a two-sample Kolmogorov Smirnov test (**ksmirnov**) or (**kstest2**) rejects that the two samples come from the same distribution.
- Compute and plot separate ECDF for 4 and 5 bidder auctions. Compare the truth x_i to the estimates \tilde{x}_i .

Part 2: LOESS

I have provided a 3% sample of the PUMS dataset – it looks at households in the US from the Census and American Community Survey. I have attached some variables like time of departure/arrival at work, transit time, whether employees receive employer sponsored health insurance or not.

- Pick one of the above variables or construct your own (such as total time at work including commuting time) and make a scatter plot of that against **totinc**
- Fit a regression that is: a) linear, b) quadratic, c) cubic in income to your variable y .
- Plot your preferred specification from the above against the scatter plot.
- Now use **lowess** and construct that plot. How if at all do the two plots differ?
- Describe in words your findings. What are we assuming about the population when we run a linear regression? A quadratic regression? A lowess regression?