Prof. Francesco Decarolis

Empirical Industrial Organization and Market Design

Problem set 3: Nonparametric Estimation of First Price Procurement Auction

The problem set has 2 main components:

1. Dataset construction
2. Estimation

For the construction of the dataset, follow the detailed instructions *DoT_instructions.pdf*.
- YOU CAN START DIRECTLY FROM "STEP 3" USING THE PDFs IN THE FOLDER
- YOU GET EXTRA POINTS FOR DOING STEPS 1 AND 2 TOO

Once you have the dataset, I want you to estimate firm cost distributions in a way similar to the example file enclosed to this problem set. If you run the jupyter notebook file *GPVprocurement.ipynb*, you will see that it produces estimates for the firms' cost distribution that are specific for the different buckets of data determined by the number of bidders in the auctions of the example dataset (*nlp.dat*). I want you to produce analogous estimates, but for the true data that you downloaded from the web. In particular, I want you to send to the class TA by the deadline of this assignment a pdf document where you report one plot containing only three step functions that represent your estimates of the bidders' cost distribution obtained for three different levels of competition: low, medium, high. So, while in the current code the plot is done for all different values that the variable number of bidders can take, I ask you to just select 3 different values in your dataset that are the 25th, 50th and 75th percentiles of the number of bidders.[1]
I suggest that you to first familiarize with the *GPVprocurement.ipynb* code, and only afterwords you prepare your code to perform the estimation. In the file *nlp.dat* you have the winning bid of 144 first price sealed bid auctions that a government used to procure some construction works. In the first column of the file there is an auction identifier (t), in the second there is the number of bidders (Nt), in the third there is the winning bid (Wt) and in the fourth a covariate (Zt). The winning bid that you observe is the lowest price offered to complete the work (i.e. the government wants to minimize its cost). Assume that: (a) there is no binding reserve price; (b) for every bidder the cost of performing the job, C, is an independent and identically distributed draw from a distribution $F_C(c)$ which is common knowledge to all bidders (i.e. assume the symmetric IPV model); (c) bidders are risk neutral and behave noncoperatively
Given these assumptions, you know that the symmetric Bayes-Nash equilibrium bid function for a generic bidder at the auction is as follows. Without loss of generality we can focus on

[1] If in your dataset there are too few observations corresponding to the exact number of bidders at which these three percentiles are located, feel free to pool together close by auctions (i.e., if the 50th percentile is 7 bidders but you have only 10 auctions with exactly 7 bidders), then impose that also auctions with slightly more/less than 7 bidders are pooled together with the auctions with 7 bidders.

bidder 1. His objective is to maximize his expected profits:

$$(s_1 - c_1)\{1 - F_C[\hat{\sigma}^{-1}(s_1)]\}^{(N-1)}$$

The necessary first order condition is:

$$1 - F_C[\hat{\sigma}^{-1}(s_1)] = (s_1 - c_1)(N-1) f_C[\hat{\sigma}^{-1}(s_1)] \frac{d\hat{\sigma}^{-1}(s_1)}{ds_1}$$

Since at the BNE $\hat{\sigma}^{-1}(s) = c$ and $\frac{d\hat{\sigma}^{-1}(s_1)}{ds_1} = \frac{1}{\sigma'(c)}$ the FCOC can be written as:

$$\sigma'(c) = \frac{[\sigma(c) - c](N-1) f_C(c)}{1 - F_C(c)}$$

The solution to the above differential equation is the BNE bid function:

$$\sigma(c) = c + \frac{\int_c^\infty [1 - F_C(u)]^{(N-1)} du}{[1 - F_C(c)]^{(N-1)}}$$

Ignoring the covariate Zt, the attached Jupyter Notebook file estimates nonparametrically the cdf of the bidders' cost: $F_C(c)$ for every realization of N. See the attached jupyter notebook file, and remember that an estimator for $F_S(s)$ is the empirical analog. The file also estimates nonparametrically the pdf of the bidders' cost: $f_C(c)$ for every realization of N. A kernel-smoothing method is needed to estimate $f_C(c)$, see Guerre, Perrigne, Vuong (2000) for the details on the threewight kernel-somoother used.
I want you to obtain these same objects for the dataset that you constructed and for the three levels of competition mentioned above (25th, 50th and 75th percentiles of the number of bidders). I want you to report these plots in a pdf file and comment (briefly) on your findings.