Advanced Industrial Organization II
Problem Set 3

DUE: Sunday Mar. 13th at noon

Please typeset your assignment in Latex. For the coding part feel free to use your preferred programming language. Please comment your code clearly, but not excessively, and use low-levels commands. The clarity of your code and the overall assignment will be evaluated. Submit your assignment through Canvas in a single zip file. You can work in a group of two and submit a single file per group.

IMPORTANT: You are required to solve ONLY ONE of the two exercises.

Exercise 1

The file ps3_ex1.csv contains data on a dynamic entry game played by $N = 5$ firms in $M = 400$ independent markets. For each firm i in market m we observe the entry/exit decision in two subsequent periods $t = 0$ and $t = 1$ and the size of each market at the beginning of period $t = 1$. We denote by $a_{imt} \in \{0, 1\}$ the status of firm i in market m in period $t \in \{0, 1\}$ and by s_{mt} the size of the market in period t which we assume can take values on a discrete support $S = \{1, 2, 3, 4, 5\}$. As typically done in the literature we will assume that the data for each market are generated from the same Markov-Perfect Equilibrium (MPE).

The flow profit of firm i in market m and period t if it decides to be active is given by:

$$\begin{aligned}\pi(s_{mt}, a_{imt} = 1, a_{-imt}, a_{imt-1}; \theta) = \theta_s \log(s_{mt}) - \theta_n \log\left(1 + \sum_{j \neq i} a_{jmt}\right) \\ - \theta_{f,i} - \theta_e(1 - a_{im,t-1}) + \epsilon_{imt}(1)\end{aligned}$$

whereas the profits from being inactive is simply $\pi(s_{mt}, a_{imt} = 0, a_{-imt}, a_{imt-1}; \theta) = \epsilon_{imt}(0)$. The shocks are assumed to be $\epsilon_{imt}(a_{it}) \sim i.i.d T1EV(0, 1)$ and private information of firms while the rest of the variables are common knowledge. Moreover, assume that in each market m the market size s_{mt} evolves over time with the following transition matrix:

$$F_s = \begin{bmatrix} 0.8 & 0.2 & 0.0 & 0.0 & 0.0 \\ 0.2 & 0.6 & 0.2 & 0.0 & 0.0 \\ 0.0 & 0.2 & 0.6 & 0.2 & 0.0 \\ 0.0 & 0.0 & 0.2 & 0.6 & 0.2 \\ 0.0 & 0.0 & 0.0 & 0.2 & 0.8 \end{bmatrix}$$

Finally, even though we only observe 2 periods in the data, the game is an infinite horizon game in which firms in each period decide to be active or not simultaneously. We will use variation across markets to estimate the parameter vector θ.

1. What are the common knowledge state variables? What's is the cardinality of the common knowledge state space?

2. Let P be a vector of CCPs and $V_i^P(x)$ the value function of firm i at a given observed common knowledge state vector x. Write the integrated Bellman equation for $V_i^P(x)$. (Feel free to omit any market and time index as markets are independent and we'll be looking at stationary markovian strategies.)

3. Define a Markov Perfect Equilibrium (MPE) in terms of a fixed point in the CCPs space e.g., $P = \Psi(P, \theta)$.

4. Using the properties of T1EV, show that

$$E\left[\epsilon_i(a_i)\middle| x, \arg\max_a\{v_i^P(a,x) + \epsilon_i(a)\} = a_i\right] = \gamma - \log(P_i(a_i|x))$$

   where $v_i^P(a,x)$ denotes the choice-specific value function e.g., s.t. $V_i^P(x) = \int \max_a\{v_i^P(a,x)+\epsilon_i(a)\}g(\epsilon_i)d\epsilon_i$ and $\gamma \approx 0.57$ is the Euler Mascheroni constant.

5. Let P^* be a MPE, rewrite $V_i^{P^*}(x)$ in a way that allows you to solve for the $|X|$ dimensional vector $V_i^{P^*}$ as a function of P^* only and call that function $\Gamma(P^*)$.

6. Define the pseudo likelihood function as

$$Q_M(\theta, P) = \sum_{m=1}^{M}\sum_{t=1}^{T}\sum_{i=1}^{N} \log \Psi_i(a_{it}|x_t; P, \theta)$$

   implement the following 2-step estimator:

i. For each i and x estimate the probability of of being active in period $t = 1$, $P_i(a_{imt} = 1|x_{mt})$ non-parametrically (e.g., frequency estimator) and denote $\hat{P}_0$ the vector of estimated CCPs.

ii. Estimate θ using the 2-step estimator $\theta_{2S} = \arg\max Q_M(\theta, \hat{P}_0)$

7. Repeat the estimation in point 6. using a binary logit instead of the frequency estimator to estimate $P_i(a_{imt} = 1|x_{mt})$ in step i.

8. Repeat the estimation in point 6. drawing randomly $P_i(a_{imt} = 1|x_{mt})$ from an $U[0, 1]$ in step i.

9. Implement the following $K-$step nested pseudo likelihood (NPL) estimator. At the $k-$th iteration, let $\hat{P}_{k-1}$ be the current guess of the CCPs, maximize the pseudo likelihood $\hat{\theta}_k = \arg\max_\theta Q_M(\theta, \hat{P}_{k-1})$ and update the CCPs using the MPE equilibrium fixed point mapping $\hat{P}_k = \Psi(\hat{P}_{k-1}; \hat{\theta}_k)$. As convergence is not guaranteed stop at the $K = 20$ iteration unless $||\hat{P}_k - \hat{P}_{k-1}|| < \delta$ at some $k < 20$. Run the routine for each of the initial $\hat{P}_0$ guesses in point 6., 7. and 8. above. In each case report whether the NPL estimator converged or not and if it converged at which iteration it did.

10. Compare your estimates in points 6., 7., 8., and 9. Explain your findings.

Exercise 2

For this assignment, you will replicate parts of Bajari and Hortaçsu (2005)'s paper on experimental auctions. You should start by reading the first part of the paper to familiarize yourself with the methods and data the authors use. This assignment is to replicate most of the computations in Sections II through IV of the paper (more details below), as well as perform some other useful calculations. The dataset ps3 auction.csv came from three auction experiments originally analyzed by Dyer, Kagel, and Levin (1989). We have excluded the initial runs of the experiments from ps3 auctions.csv, leaving the observations that Bajari and Hortaçsu use in the paper. Before writing any code, read the data section of Bajari and Hortaçsu to learn more about the structure of the data set. Start by familiarizing yourself with the data.

1. Reproduce the two plots in Figure 1 of the paper. Compute and store the difference between the observed bids (BidC3, BidC6) and the Nash equilibrium bids.

2. Using the bids within each experiment, compute the empirical CDF for each type of bid (N = 3; N = 6). Use the evenly-spaced grid 0.01, 0.02, 0.03, ..., 29.99, 30.00 to evaluate the empirical CDF. Foreach auction size, plot of the empirical CDFs of the three experiments in the same panel. Is the distribution of bids much different across experiments?

3. Use the appropriate empirical CDF to compute the expected profit and the optimization error (defined in equations 3 and 4) for each bid.

4. Using the above calculations, reproduce Table 1. The differences from your calculations should match Table 1 to rounding error.

Next, replicate some of the computations in Section III.

1. Using Silvermans rule of thumb for the bandwidth, compute an estimate of g(b) for each bid in the data set. For a visualization of the bid data, provide a plot for these bids for each bid type.

2. For each bid in the data set, estimate the CDF of bids $G(b)$ using the empirical CDF, $\hat{G}(b)$. Do this separately for the N = 3 bids and the N = 6 bids.

   - Compute these estimates two ways as discussed in footnote 14: (1) Pool the bids across ex- periments and use this "full" data set to compute the density and CDF estimates, (2) Use the data within each experiment to compute the density and CDF estimates. In either case, use Silvermans rule of thumb for the density estimates.

3. Use equation (9) and previous calculations to compute two estimates for each valuation: One based on the N = 3 bid data and one based on the N = 6 bid data. Summarize these estimates of the valuation with two types of plots:

   - For each auction size, provide a scatterplot of your estimates of the valu- ation versus the true valuation. For consistency, put true valua- tion on the horizontal axis and plot the 45-degree line.

   - Plot the histogram of the true valuations and your two estimates of the valuations. That is, reproduce Figure 2 in the paper.

4. For each way of estimating the CDF and pdf in (2), compute the L1 and L2 norm.

For the next part of this exercise, replicate some of the computations in Section IV (Risk Aversion).

1. Compute and store the 0 through 100 integer percentiles of the bid data. At each of these percentiles, compute CDF and pdf estimates. Use these percentiles to construct the term in brackets on the right hand side of equation (15).

2. Estimate θ using equation (15) and ordinary least squares regression using (i) the full sample of percentiles, (ii) the trimmed sample containing per- centiles 5 through 95, and (iii) the trimmed sample containing percentiles 25 through 75. Report point estimates and 95 percent confidence intervals for your estimate of risk aversion.

3. Use the point estimate of risk aversion you obtained from the trimmed sample of percentiles 25 through 75 to compute two estimates for each valuation (one for N = 3; one for N = 6).

   - As for the risk neutral case, provide a scatter plot of your estimates of the valuation versus the true valuation. For consistency, put true valuation on the horizontal axis and plot the 45-degree line.
   - Plot the histogram of the true valuations and your two estimates of the valuations. That is, reproduce Figure 2 in the paper.
   - Compute the L1 and L2 norm using these new estimates for the valuation.