Advanced Industrial Organization II Problem Set 3

Ali Hortaçsu and Kyeongbae Kim*

February 19, 2019

Due on March 12, at the beginning of the class.

1 Rust (1987): Bus engine replacement problem

Harold Zurcher has one bus, which runs a middle-of-nowhere city. At the beginning of each period, he has to decide whether to replace the bus engine. When the bus engine is replaced ($i_t = 1$), the mileage becomes zero at the beginning of the period. Then the bus runs for a period, and the mileage, x_t, is recorded at the end of the period. Replacing the bus engine takes a fixed cost, RC. There is a per-period disutility as the mileage accumulates (such as an increasing per-period maintenance cost). In addition, there exists a mean-zero choice specific T1EV iid shocks. To formalize, the per-period utility has the following form:

$$u\left(x_t, i_t, \epsilon_t; \theta\right) = \begin{cases} -c\left(x_t; \theta\right) + \epsilon_{0,t} & \text{if } i_t = 0 \\ -RC - c\left(0; \theta\right) + \epsilon_{1,t} & \text{if } i_t = 1, \end{cases}$$

where $c\left(x_t; \theta\right) = \theta_1 x_t + \theta_2 x_t^2$. End-of-period mileage never decreases unless the engine is replaced at the beginning of the period. $(x_t, \epsilon_{0,t}, \epsilon_{1,t})$ follows a Markov transition probability, $p(x_{t+1}, \epsilon_{0,t+1}, \epsilon_{1,t+1} | x_t, \epsilon_{0,t}, \epsilon_{1,t}, i_t, \theta_3)$. The econometrician observes x_t. Let $\theta = (\theta_1, \theta_2, \theta_3, RC)$ be parameters of interest.

1. How do you recover the engine replacement choice, i_t, from the mileage data, `ps3_mileage.csv`?

2. What is Rust (1987)'s conditional independence assumption, which we will maintain for the rest of the problem? Discuss briefly.

3. Discretize x_t and estimate the Markov transition probability, $p(x_{t+1} | x_t, i_t, \theta_3)$. (Hint: $p(x_{t+1} | x_t, i_t = 0, \theta_3)$ is a $K \times K$ upper-triangular matrix, where each row adds up to 1. What is a dimension of

*hortacsu@uchicago.edu, kbkim@uchicago.edu

$p(x_{t+1}|x_t, i_t = 1, \theta_3)$? K and a way to discretize are subject to a researcher's choice. While not necessary, for this part you can assume $p(x_{t+1} - x_t|x_t, i_t, \theta_3) = p(x_{t+1} - x_t|i_t, \theta_3)$ if it makes your life easier.)

4. Derive $Pr(i_t = 0|x_t; \theta)$ and $Pr(i_t = 1|x_t; \theta)$, as functions of the expected value function, $EV_\theta(x, i) = \int V_\theta(y, \epsilon)p(d\epsilon)p(dy|x, i, \theta_3)$, and a per-period utility function, $\bar{u}(x_t, i_t; \theta)$, net of ϵ. (Hint: Be careful about timing x and y are observed, respectively.)

5. Show the following fixed point equation holds:

$$EV_\theta(x, i) = \int_y \log \left\{ \sum_j \exp[\bar{u}(y, j; \theta) + \beta EV_\theta(y, j)] \right\} p(dy|x, i, \theta_3).$$

6. Estimate the model parameters, θ, using Rust's full ML method, for the discount rate $\beta = 0.9$.

2 Bajari and Hortaçsu (2005)

For this assignment, you will replicate parts of Bajari and Hortaçsu (2005)'s paper on experimental auctions.[1] 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. For each auction size, plot of the empirical CDFs of the three experiments in the same panel. Is the distribution of bids much different across experiments?

[1] "Are Structural Estimates of Auction Models Reasonable? Evidence from Experimental Data." Patrick Bajari and Ali Hortaçsu, *The Journal of Political Economy,* Vol. 113, No. 4 (August 2005), pp. 703-741.

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.[2]

Next, replicate some of the computations in Section III.

1. Using Silverman's 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.

   (a) Compute these estimates two ways as discussed in footnote 14: (1) Pool the bids across experiments 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 Silverman's 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:

   (a) For each auction size, provide a scatterplot 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.

   (b) 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).[3]

[2]Note: The number of observations reported in Table 1 of the paper - $N = 414$ - does not match the number of observations in the data set we gave you - $N = 408$. The reported number of observations in the table is what is wrong here, not the data set. The data set with 408 observations should replicate Table 1 with only a few differences of 0.01 interspersed in the table.

[3]Depending on the method your computing package uses to compute percentiles, your results may dier slightly from the ones presented in the paper.

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 percentiles 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. Perform the alternative verification of the risk aversion hypothesis outlined on the bottom half of page 721.

4. 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 $N = 6$).

   (a) 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.

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

   (c) Compute the L1 and L2 norm using these new estimates for the valuation.