Problem Set 2 ECON8853

Joe Jourden, Federico Rodari, Tuo Yang

November 14, 2022

Code naming convention: part 1 (Berry 1992), part 2.1 (Ciliberto and Tamer 2009 - objective function), part 2.2 (Ciliberto and Tamer 2009 - confidence region)

1. **Berry (1992)** Let θ be the vector of all parameters, and let N_m be the simulated number of firms in market m. Then our MLE is defined as

$$
\begin{aligned}
(\hat{\mu}, \hat{\sigma}^2) &= \underset{(\mu,\sigma^2)}{\arg\max} \prod_{m=1}^{M} \Pr(N_m = N_m^* \mid X, \theta) \\
&= \underset{(\mu,\sigma^2)}{\arg\max} \sum_{m=1}^{M} \ln(\Pr(N_m = N_m^* \mid X, \theta)).
\end{aligned}
$$

   The resulting estimates are $\widehat{\mu} = 2.178$, $\widehat{\sigma}^2 = 0.940$. Note that here we can only estimate the probability of observing a number of entrants without making any conclusion on the identity of the entrants. By using the order of entry based on the profitability of a firm we can pin down a unique number of entrants in order to avoid the coherency issue arising from multiple equilibria.

2. **Ciliberto and Tamer (2009)** Here we follow the paper and the provided supplementary material. In other words, we compute the bounds with simulation and create the confidence region by iterating the subsampling procedure proposed by CHT. To check the equilibrium condition, we rely on the definition that for a given market configuration y^j we must observe that the firms entering make non-negative profits and those who stay out would make negative profit by entering. Three main comments:

   - To estimate the conditional probability of a given market structure we used a multinomial logit (estimated in Stata) and attached probability 0 to the outcomes that were not observed in the provided dataset.
   - The only difference with respect to the original paper is that we are using i.i.d draws while the paper uses dependent draws by multiplying the cholesky factor of the covariance matrix of the simulation ϵ^r by the simulated vector.
   - In our true world μ is a parameter, but we are computing the confidence region for a set based on the pseudocode. If we want to follow the approach to construct the confidence region of a point estimate, we should follow the alternative procedure provided in the supplement.

   Our estimated bounds (we iterate c_i two times as in the paper) are $\mu_{lb} = 1.975$, $\mu_{ub} = 2.8$. Please note that we included the random seed only after getting these results, so running the code might give different results (the code took more than 2 hours to run so for time's sake we did not run it again given the time).