ECON 761, Fall 2022 - Problem Set 5

Due before the end of the day, on Sunday 11/14.

Download the data from Jia's (2008) Econometrica paper at this location. Upload the file containing the 1997 data into your favorite statistical package.

1. Run Probit regressions of entry by Wal-Mart and entry by Kmart on market level characteristics, including the presence of the competing firm. Find, for each firm, the specifications that seem to best fit the data. Report your estimates in Table I, and comment briefly. What is the biggest weakness of this model?

2. Consider an instrumenting strategy (control function?) that would use excluded variables as instruments in the Probit regressions above. In particular, consider the Probit regression of Kmart entry on covariates and Wal-Mart entry. Consider the variable "distance from Benton county"; could it be used as an instrument for Wal-Mart entry? (Hint: think of strategic nature of the interaction that generates the data...)

3. Now consider a "Bresnahan and Reiss" analysis of this industry. Estimate an ordered Probit model in two specifications, where the dependent variable is (i) number of large players (Wal-Mart and Kmart), and (ii) total number of small and large players. Use as market-level covariates the same variables you used in 1. Report your estimates in Table II, and compare with the results you found in Table I. What are the main limitations of this model?

4. To add heterogeneity to the model without having to confront the issue of multiple equilibria, suppose Wal-Mart and Kmart play an entry game of incomplete information (this is the same assumption maintained in Seim (2006)). The model specifies profit functions for firm i in market m:

$$\pi_{im}(y) = y_{im}\left(x_{im}\beta_i + \Delta_i y_{-i} + \varepsilon_i\right),$$

   where β_i and Δ_i are parameters, x_{im} are firm-market level observables, and ε_i is the unobservable part of profits. To estimate this game, use the two-step method proposed by Bajari et al. (2012). To keep things simple, the "first step" can be fully parametric.[1] In contrast to Jia (2008), do not consider the presence of chain-effects, and suppose that the two players enter each market with separate simultaneous decisions. Report your estimates in Table III, and briefly comment. There is no need to provide correct standard errors, but outline briefly how you could compute those.

5. [**Extra Credit**] Consider the same model outlined in part 4., but change the estimation strategy. You have now to obtain coefficient estimates for the game use the MPEC approach outlined in Su (2014), starting from the AMPL code he provided (in the Supplementary Materials folder). AMPL is a very useful modeling language that allows to write down optimization programs in a language very close to math, and efficiently use state-of-the art solvers by providing automatic differentiation. For a general introduction to AMPL, read this book. To estimate this model with AMPL, you need to:

   (a) Obtain a `.dat` file which contains the data you'll use in the analysis in the form required by AMPL. A preliminary consideration about data is that Su's method requires panel-like data, i.e. multiple observations for the same market. To generate a similar feature in our dataset, which is cross sectional, the space of covariates can be discretized, so that we have multiple entry observations for the same "discretized" markets. You can do this yourself, for instance in Matlab, by dividing into quantiles the continuous covariates, and substituting x_t with $\overline{x}^q$ such that $x_t \in [x_{q-1}, x_q]$. Once you have done this, you can create a .dat file with Matlab using the function `fprintAmplParamCLSU.m` that I uploaded on the website. Alternatively, you can use the file `DataSU.dat` that I provide. This file contains the following variables:

[1] Alternatively, you could discretize of covariates - read point (a) below for guidance.

i. `nM`: number of discretized markets (80),
ii. `nN`: actual number of markets (2065),
iii. `nV`: number of market level variables (5),
iv. `XX`: matrix `nM`×`nV` of covariates, containing the variables # of small stores, log of county population, dummy for Midwest, dummy for South, constant,
v. `x`: `nN` vector that assigns each market to its discretized counterpart,
vi. `y`: matrix `nN`×2 that indicates entry by Walmart and Kmart, respectively,

(b) Write a `.mod` file that outlines the problem you want to solve. Only slight modifications of the file provided by Su are necessary, but make sure you understand the logic of that file well.

(c) Write a `.run` file that collects the model `.mod` and the data file `.dat`, and instructs the solver to provide a solution. To obtain estimates, you can use the facilities of the NEOS Server, a free internet-based service for solving optimization problems, hosted by the Institute of Discovery here at UW-Madison. Make sure your files follow the instructions outlined here.

Present your estimates in table IV, and comment on similarities and differences with the other results. There is no need to provide correct standard errors, but outline briefly how you could compute those.

References

[1] Bajari, Patrick, et al. "Estimating static models of strategic interactions." Journal of Business & Economic Statistics (2012).

[2] Jia, Panle. "What happens when Wal-Mart comes to town: An empirical analysis of the discount retailing industry." Econometrica 76.6 (2008): 1263-1316.

[3] Seim, Katja. "An empirical model of firm entry with endogenous product-type choices." The RAND Journal of Economics 37.3 (2006): 619-640.

[4] Su, Che-Lin. "Estimating discrete-choice games of incomplete information: Simple static examples." Quantitative Marketing and Economics 12.2 (2014): 167-207.