Economic and Statistics III

Problem Set 4

In this problem set we will apply some of the procedures for modeling dynamic discrete choice models. Specially, we will consider empirical procedures for distinguishing between two similar but distinct models (recall 1[w] is the indicator function equal to 1 if w is true and 0 otherwise) :

$$
\begin{aligned}
&d_{it} = 1\left[z_{it}\beta + \alpha_i + \varepsilon_{it} \ge 0\right], \quad i = 1,\dots,N; \quad t = 1,\dots,T \\
(1) \qquad &\alpha_i \perp \varepsilon_{it}, \quad \alpha_i \sim iid\ N\left(0,\sigma_\alpha^2\right), \quad \varepsilon_{it} \sim iid\ N\left(0,1\right) \\
&E\left[\varepsilon_{it} \mid z_{i1},\dots,z_{iT},\alpha_i\right] = 0, \quad E\left[\alpha_i \mid z_{i1},\dots,z_{iT}\right] = 0
\end{aligned}
$$

$$
\begin{aligned}
&d_{it} = 1\left[z_{it}\beta + \gamma d_{it-1} + \varepsilon_{it} \ge 0\right], \quad i = 1,\dots,N; \quad t = 1,\dots,T \\
(2) \qquad &E\left[\varepsilon_{it} \mid z_{i1},\dots,z_{iT},d_{it-1}\right] = 0, \quad \varepsilon_{it} \sim iid\ N(0,1)
\end{aligned}
$$

Task one concentrates on applying the procedures; task two requires an understanding of the conceptual framework linking these procedures. Download from the class web page the (Windows) executable program, e719.exe. The program will prompt you for the name of an output file, and a seed value. Depending on the value of the seed, the program will generate a dataset following either model 1 or model 2. The data are written in the current directory in the output file name. The sample data contain observations on 1,000 cross sectional units (N=1000), each observed 8 periods (T=8). The sample window observes the dynamic process starting in period 25 (so we do not have observations from the beginning of the decision making process). The person-period records (it) are structured as (i, t, d_{it}, z_{it}).

Task 1. Run e719.exe twice, with two different seeds (and different file names). You must report the seed values used to generate the data when you turn in your assignment. For each of the two data sets, determine whether model (1) or model (2) generated the data. Report preferred estimates (and standard errors etc.) for the parameters of each model, and *diagnostic information in support of your conclusion.* Describe the statistical properties of your estimates and specification testing procedures.

You may use any statistical package of your choosing. I did the assignment using standard routines in STATA. You are not limited to procedures described in class. This is an empirical exercise and you should feel free to make liberal use of results in Wooldridge or other textbooks or articles.

Task 2. Consider the dynamic discrete choice model:

$$
\begin{aligned}
&d_{it} = 1\left[z_{it}\beta + \gamma d_{it-1} + u_{it} \ge 0\right], \quad i = 1,\dots,N; \quad t = 1,\dots,T \\
(3) \qquad &u_{it} = \rho u_{it-1} + \varepsilon_{it}, \quad \varepsilon_{it} \sim iid\ N\left(0,1\right) \\
&E\left[\varepsilon_{it} \mid z_{i1},\dots,z_{iT},d_{it-1}\right] = 0
\end{aligned}
$$

(a) Assume you observe the process from the beginning; $d_{io} = d$. Is the GMM estimator of Avery-Hansen-Hotz consistent? Why or why not?
(b) Write out the likelihood function.
(c) Why is model (3) so much more computationally demanding than models (1) and (2) of Task 1. Would it be less demanding if ε_{it} were a logistic random variable? Why?
(d) Now assume that you do observe the process beginning in period τ >0. Modify the likelihood function and discuss different computation and estimation strategies.