Structural Estimation Methods Fall 2021
Assignment: Single Agent Dynamic Programming/Estimation Due Dec 15

Consider a capital replacement problem similar to that in Rust (1987). Firms each use one machine to produce output in each period. These machines age, becoming more likely to breakdown, and in each time period the firms have the option of replacing the machines. Let a_t be the age of your machine at time t and let the expected current period profits from using a machine of age a_t be given by:

$$\Pi(a_t, i_t, \epsilon_{0t}, \epsilon_{1t}) = \begin{cases} \theta_1 a_t + \epsilon_{0t} & \text{if } i_t = 0 \\ R + \epsilon_{1t} & \text{if } i_t = 1 \end{cases}$$

where $i_t = 1$ if the firm decides to replace the machine at t, R is the net cost of a new machine, and the ϵ_t's are time specific shocks to the utilities from replacing and not replacing. Assume that these ϵ_t's are i.i.d. logit errors.

Lets assume a very simple state evolution equation:

$$a_{t+1} = \begin{cases} \min\{5, a_t + 1\} & \text{if } i_t = 0 \\ 1 & \text{if } i_t = 1 \end{cases}$$

In words, if the firm decides not to replace, the machine ages by one year (up to a maximum of 5 years - after 5 years machines don't age). If the firm replaces in the current year, the age next year is 1. Note that there are thus only 5 possible values of a_t - 1,2,3,4, and 5.

1) Write down the dynamic programming problem for a firm maximizing the PDV of future profits (assume an ∞ horizon).

2) What are the differences between this $\Pi(a_t, i_t, \epsilon_{0t}, \epsilon_{1t})$ and the profit function in the class notes on Rust? What happened to $c(0; \theta)$?

3) On the computer, write a procedure that solves this dynamic programming problem given values of the parameters (θ_1, R). Assume that $\beta = .9$. Use the "alternative-specific" value function method from in class, i.e. define $\overline{V}_0(a_t)$ and $\overline{V}_1(a_t)$ - you should end up with equations looking something like

$$\begin{aligned} \overline{V}_0(a_t) &= \theta_1 a_t + \beta E\left[\max\left\{\overline{V}_0(a_{t+1}) + \epsilon_{0t+1}, \overline{V}_1(a_{t+1}) + \epsilon_{1t+1}\right\}\right] \\ \overline{V}_1(a_t) &= R + \beta E\left[\max\left\{\overline{V}_0(a_{t+1}) + \epsilon_{0t+1}, \overline{V}_1(a_{t+1}) + \epsilon_{1t+1}\right\}\right] \end{aligned}$$

and do the contraction mapping on these two 5-vectors. Iterate the contraction mapping until the $\overline{V}$ functions dont change very much. Remember that given the logit error assumption there is an analytic solution to the expectation of the max in these equations (and Euler's constant $\approx$.5772).

4) Solve the model for the parameters $(\theta_1 = -1, R = -3)$. Suppose $a_t = 2$. Will the firm replace its machine in period t? Oops, that was a trick question - for what value of $\epsilon_{0t} - \epsilon_{1t}$ is the firm indifferent between replacing its machine or not? What is the probability (to an econometrician who doesn't observe the ϵ's) that this firm will replace its machine? What is the PDV of future profits for a firm at state $\{a_t = 4, \epsilon_{0t} = 1, \epsilon_{1t} = -1.5\}$? (the constant term has been normalized so it is OK if this PDV is <0)

5) There is a simulated dataset on Canvas, "data.asc". This dataset has just two columns - a_t (first column) and i_t (second column). Consider this as cross-sectional data - i.e. there is only one data point per firm. Estimate (θ_1, R) using maximum likelihood. The estimates should end up being approximately $\theta_1 = -1.1, R = -4.4$. Your ML function evaluation should look something like this:

a) Start with arbitrary (θ_1, R)

b) Solve the dynamic programming problem given these parameters (i.e. compute the functions $\overline{V}_0(a_t)$ and $\overline{V}_1(a_t)$)

c) Using $\overline{V}_0(a_t)$ and $\overline{V}_1(a_t)$, compute the probability of replacement for each a_t, i.e. $\text{Prob}(i_t = 1 \mid a_t)$

d) Compute the likelihood of each observation j, e.g. $L_j = i_j$ Prob$(i_j = 1 \mid a_j) + (1 - i_j)(1-$Prob$(i_j = 1 \mid a_j))$

e) Return $-\ln(L) = -\sum_j \ln(L_j)$ (the minus is if you are using a minimization (rather than maximization) procedure)

Explain why you think I've suggested estimation is based on a conditional likelihood (conditional on a_t).

6) Compute standard errors of your estimates (you should use the regular maximum likelihood standard error formulas).

7) Describe (you do NOT have to do this on the computer) how you would need to change your model (either the dynamic programming problem, the estimation procedure, or both) to accommodate the following perturbations of the model:

a) Consider an alternative empirical model. Suppose there are two types of firms - differing in their value of θ_1. Proportion α of firms have $\theta_1 = \theta_{1A}$, proportion (1-α) have $\theta_1 = \theta_{1B}$. How would you change both the dynamic programming problem and the likelihood function? You can ignore the potential initial conditions problem in the likelihood function.

b) What if you had the model in a) but you have panel data, i.e. multiple observations on each firm? Write down the likelihood function (again, ignoring potential initial conditions problems).

c) Continuing with the panel data situation, what if instead of *firms* differing in θ_1, *machines* differ in θ_1, i.e. when a firm replaces their old machine, the new machine may have $\theta_1 = \theta_{1A}$ (w/prob α) or it may have $\theta_1 = \theta_{1B}$ (w/prob 1-α)? Again, describe how the dynamic programming problem and likelihoods change (ignoring potential initial conditions problems)

d) Describe the initial conditions problem that you have ignored in the prior three parts. Propose at least one way to address this initial conditions problem.

e) Go back to only one "type" of firm and one observation per firm. What if a_t does not evolve deterministically, i.e. if you don't replace ($i_t = 0$):

$$a_{t+1} = \begin{cases} a_t & \text{with probability } \lambda \\ \min\{5, a_t + 1\} & \text{with probability 1-}\lambda \end{cases}$$

Make an informal argument that there is some information in the data regarding the parameter λ. Note that this is not obvious because you only have 1 data point per firm and thus you never actually observe transitions from a_t to a_{t+1}. You can assume that your data is a random sample of firms that are in a "steady state" (i.e. they have existed for a long time).

8) Estimate the original model using the Hotz and Miller algorithm. Recall that the first thing you need to do here is compute "non-parametric" estimates of the replacement probabilities in the data (i.e. $\widehat{P}(\cdot)$ from the class notes). Compare your estimates to those in part 4). You do not need to compute standard errors.

9) **Optional:** iterate this estimation procedure (as described in the class notes) and see how many iterations it takes to get "close" to the ML estimator (AM shows that this should eventually converge to the ML estimator).

10) **Optional:** Try this again, where you do not start with $\widehat{P}(\cdot)$ estimated from the data, but instead use an arbitrary (i.e. just take a guess) set of replacement probabilities at each state instead. The Hotz-Miller algorithm based on this arbitrary $\widehat{P}(\cdot)$ will not produce consistent estimates, but according to AM repeatedly iterating the estimator until convergence should. Observe how many iterations are necessary to get close to the original ML estimator, and if you want, how much this depends on how wrong your initial replacement probabilities are.