Paul Le Tran
plt377
14/12/21
ECO388E

Problem Set 3

1).

Observe that the state variables needed to define the current period's profits and expectations about discounted future profits are a_t, ϵ_t. This means we can write the firm's value function as follows:

$$V(a_t, \epsilon_t; \theta) = max_{i_t}\left(E\left[\sum_{j=t}^{\infty} \beta^{j-t}\, \Pi\left(a_j, i_j, \epsilon_j, ; \theta\right)\middle| a_t, i_t; \theta\right]\right).$$

Converting this into a Bellman equation yields the following:

$$V(a_t, \epsilon_{0t}, \epsilon_{1t}; \theta) = max_{i_t}(\Pi(a_t, \epsilon_{0t}, \epsilon_{1t}; \theta) + \beta E[V(a_{t+1}, \epsilon_{0t+1}, \epsilon_{1t+1}; \theta)|a_t, i_t; \theta]).$$

2).

The first major difference between the problem set and Rust (1987)'s original problem (i.e., the HZ problem) is that in the latter, the bus engine mileage, x_t, is a continuous variable. In the former, the equivalent machine age, a_t, is a discrete variable that can only take values in the set {1, 2, 3, 4, 5}. The second difference we see between the profit functions of the problem set and the class notes is that the expected cost function of operating a machine with $(a\_\{t\}, c(a\_\{t\};\ \theta)$, is equal to $\theta_1 a_t$ in the problem set. As a result, we have in the problem set that $c(0; \theta) = 0$. This makes sense because having an age of $a_t = 0$ is not technically possible.

Other technical differences are that the signs of R and $c(a_t; \theta)$ are reversed between the two profit functions. However, this should only be notational differences, not actual-value differences. This is because both profit functions still treat the replacement cost, R, and expected cost function, $c(a_t; \theta)$, to be negative.

3).

The following is the code set up to perform Rust (1987)'s "alternative-specific" value function iteration:

```
%==========================================================================
% Initialising discounting parameter and Euler's constant
beta = 0.9;
gamma = 0.5772;

% Creating vector that will house all possible values of a_{t}. Recall that
% once a machine reaches age 5, it will stay at age 5 forever until
% replaced.
a = (1:5)';

% Creating initial values for both value functions. For simplicity, we are
% making the value of not replacing or replacing, in all ages, to equal one
% initially.
v0 = ones(length(a), 1);
v1 = ones(length(a), 1);

%======================================================================
% Problem 4 code (setting parameter values to actually do the VFI loop)
%======================================================================
% Creating initial value for parameters
%=====
% NOTE
%=====
% Paramater value mapping:
% theta(1) = \theta_{1}
% theta(2) = R;
%=========
% END NOTE
%=========
theta0 = [-1 -3];

% Creating variable that stores max(abs(v - v_next)), the error metric. We
% are focusing on the max of the error because if the max value satisfies
% the threshold, every other value of the vector will.
error = 1;

% Creating variable that stores iteration number
iteration = 0;

%===============================
% Value function iteration loop
%===============================
% Iterating to find fixed point of value function. The threshold that stops
% the loop will be if the max error of the individual errors for v0 and v1
% is <= 0.001.
while error >= 0.001
  % Displaying iteration number
  disp(iteration);

  % Creating next-period age value function vectors that go into the RHS of
  % the value functions created by using initial or previous value
  % functions
  v0a1(1: length(a) - 1, 1) = v0(2:length(a), 1);
  v0a1(length(a), 1) = v0(length(a), 1);
```

```
  v1a1(1: length(a) - 1, 1) = v1(2:length(a), 1);
  v1a1(length(a), 1) = v1(length(a), 1);

  % Creating value functions created by using initial or previous value
  % functions
  %=====
  % NOTE
  %=====
  % Paramater value mapping:
  % theta(1) = \theta_{1}
  % theta(2) = R;
  %=========
  % END NOTE
  %=========
  v0_next = @(theta) theta(1).*a + beta.*(gamma + log(exp(v0a1) +
exp(v1a1)));
  % Using repmat so the output of function v1_next is the same dimension as
  % vector of ages a.
  v1_next = @(theta) theta(2) + beta.*(gamma + log(exp(repmat(v0(1, 1),
length(a), 1)) + exp(repmat(v1(1, 1), length(a), 1))));

  % Storing max(abs(v - v_next)). Choosing the max so the abs vector has
  % every value to be below error threshold.
  error = max(max(abs(v0 - v0_next(theta0))), max(abs(v1 -
v1_next(theta0))));
  disp(error);

  % Storing previous value function
  v0 = v0_next(theta0);
  v1 = v1_next(theta0);

  iteration = iteration + 1;
end
clear iteration error v0a1 v1a1;
%==========================================================================
```

4).

Given our calculated difference in the error terms, we see that a firm is indifferent between replacing and keeping a machine whose age is 2 when said difference is equal to -0.1145. If the difference is less than -0.1145, then the firm will not replace the machine. If the difference is greater than -0.1145, then the firm will replace the machine.

Using the extreme value distribution to calculate the probability of seeing this difference (which results in seeing the firm replace its machine at age 2), the probability is 0.5286.

The PDF of future profits for a firm at state $(a_t = 4, \epsilon_{0t} = 1, \epsilon_{1t} = -1.5)$ is -14.7548. As a result, we see that it is still cheaper for the firm to replace the machine in this period than to wait until age 5 to replace.

5).

Our MLE estimation for parameters (θ_1, R), $(\widehat{\theta_1}, \hat{R})$, is (-1.1484, -4.4464).

Additionally, we are estimating based on the likelihood of i_t being conditional on a_t because the firm's decision to replace the machine depends on how old the machine is. Specifically, the firm's decision is based on the expected future costs of its machine. This in turn is dependent on the current age of the machine. As a result, the likelihood must be conditioned on a_t in order to do our estimation of parameters.

6).

The SEs for our MLE estimation of parameters (θ_1, R), $\left(\left(SE(\widehat{\theta_1}), SE(\hat{R})\right)\right)$, is (0.0749, 0.3210).

7a).

We now assume that two types of firms exist, where they differ in their value of θ_1. Furthermore, we assume that proportion α of firms have $\theta_1 = \theta_{1A}$ and proportion $(1-\alpha)$ have $\theta_1 = \theta_{1B}$. In order to accommodate these changes, we need to modify our dynamic programme problem – specifically, the Bellman equation – as follows:

$$V_j(a_{jt}, \epsilon_{0jt}, \epsilon_{1jt}; \theta_j) = max_{i_{jt}}\left(\Pi(a_{jt}, \epsilon_{0jt}, \epsilon_{1jt}; \theta_j) + \beta E\left[V(a_{jt+1}, \epsilon_{0jt+1}, \epsilon_{1jt+1}; \theta_j) \middle| a_{jt}, i_{jt}; \theta_j\right]\right).$$

Observe that the value function, replacement choice shocks, replacement decision, machine age, and parameters now have a subscript j. This is to represent that two types of firms exist now. Essentially, this means that our dynamic programming problem now involves calculating two different value functions via value function iteration.

Changes that need to be applied to the likelihood function are as follows:

$$p(i_j = 1 | a_{jt}, \epsilon_{jt}) = \alpha * p(i_j = 1 | a_{jt}, \epsilon_{jt}; \theta_{1A}) + (1-\alpha) * p(i_j = 1 | a_{jt}, \epsilon_{jt}; \theta_{1B}).$$

If we assume both the error terms are iid logit errors and conditional independence, we obtain the expression for the likelihood function:

$$p(i_j = 1 | a_{jt}, \epsilon_{jt}) = \alpha * \frac{e^{V_{1A}(a_{jt}, \epsilon_{jt})}}{e^{V_{1A}(a_{jt}, \epsilon_{jt})} + e^{V_{0A}(a_{jt}, \epsilon_{jt})}} + (1-\alpha) * \frac{e^{V_{1B}(a_{jt}, \epsilon_{jt})}}{e^{V_{1B}(a_{jt}, \epsilon_{jt})} + e^{V_{0B}(a_{jt}, \epsilon_{jt})}}.$$

7b).

Given our model writing in part 7a, if we had panel data that ends at time period T, we must now consider the sequence of decisions by a firm, $\left[i\left(a_{jt}, \epsilon_{jt}\right) = 1\right]_{t<T}$. From this, we can write the likelihood function as follows:

$$p\left(\left[i\left(a_{jt}, \epsilon_{jt}\right) = 1\right]_{t<T}\right) = \alpha * p\left(\left[i\left(a_{jt}, \epsilon_{jt}\right) = 1\right]_{t<T}; \theta_{1A}\right) + (1-\alpha) * p\left(\left[i\left(a_{jt}, \epsilon_{jt}\right) = 1\right]_{t<T}; \theta_{1B}\right).$$

As before, if we assume both the error terms are iid logit errors and conditional independence, we obtain the following expression for the likelihood function:

$$p\left(\left[i\left(a_{jt}, \epsilon_{jt}\right) = 1\right]_{t<T}\right) = \alpha * \prod_{t<T} \frac{e^{V_{1A}(a_{jt}, \epsilon_{jt})}}{e^{V_{1A}(a_{jt}, \epsilon_{jt})} + e^{V_{0A}(a_{jt}, \epsilon_{jt})}} + (1-\alpha) * \prod_{t<T} \frac{e^{V_{1B}(a_{jt}, \epsilon_{jt})}}{e^{V_{1B}(a_{jt}, \epsilon_{jt})} + e^{V_{0B}(a_{jt}, \epsilon_{jt})}}.$$

7c).

Given all the assumptions made in parts 7a and 7b, if we additionally assume that machines now can differ in parameter θ_1 in the same split as firms do (i.e., a firm's new machine could have θ_{1A} with probability α or θ_{1B} with probability $(1-\alpha)$), this is essentially saying that the model now considers machine type to be a state variable that has to be included. Additionally, machines differing will now result in our error terms ϵ_{jt} to be serially correlated. This will result in the expectations operator seen in all our "alternative-specific" value functions to be conditioned on ϵ_{jt} now. Due to there being many "alternative-specific" value functions now with the different types assumed, we will write the "general" change that all these value functions experience. More specifically, we have:

$$\overline{V_{0J}} = u\left(a_{jt}, 0; \theta_j\right) + \beta * E\left[V\left(a_{jt+1}, \epsilon_{jt+1}; \theta_j\right) + a_{jt}, \epsilon_{jt}, i_{jt} = 0; \theta_j\right];$$

$$\overline{V_{1J}} = u\left(a_{jt}, 1; \theta_j\right) + \beta * E\left[V\left(a_{jt+1}, \epsilon_{jt+1}; \theta_j\right) + a_{jt}, \epsilon_{jt}, i_{jt} = 1; \theta_j\right].$$

With regards to changes in our likelihood calculation, recall that machines differing in type translates to the choice-specific error terms to now be serially correlated to each other. As a result, the formula used to calculate the likelihood (as done in problems 3 – 6 and parts 7a and 7b) is no longer applicable. In terms of class notes, this is because assumption C is violated.

7d).

In simple terms, the initial conditions problem that was ignored in parts 7a – 7c essentially comes from the fact that we do not see the initial period of data. In other words, whilst we have a "complete" model of how $(i_{j1}, \dots, i_{jT})$ and $(a_{j2}, \dots, a_{jT})$ are determined, we do not have a "complete" model of how age a_{j1} is determined. This is because we do not observe $a_{j0}, \epsilon_{j0}, i_{0t}$. As a result, any serial correlation in the error terms that our model observes will either be due to unobserved heterogeneity or by the initial draw of ϵ_{jt}. We should note that this is a problem just for the econometrician.

One solution to the initial conditions problem is to condition on a_{j1}, which yields the following likelihood:

$$p\left(a_{j2}, \dots, a_{jT}, i_{j1}, \dots, i_{jT} \middle| a_{j1}; \theta_j\right).$$

Doing this will require us to know the density

$$p(\epsilon_1 | a_1; \theta).$$

The solution would then require simulating the above conditional distribution given the solution of the dynamic programming problem.

An additional potential solution could be to simulate ϵ_{jt} all the way back to the initial period of data.

7e).

Assume the evolution of a_t is now the following if you don't replace:

$a_{t+1} = min(5, a_t + 1)$, with probability λ;

a_t, with probability $(1 - \lambda)$.

It should be noted that if a firm does replace, the machine's age will still always be one. This all means that a machine's age isn't guaranteed to increase by one if it is kept by the firm into the next period. We also assume that our data is a random sample of firms that have existed for a long time. Despite the stochastic nature of a machine's age if it isn't replaced, there is still some information in the data that can be used to figure out \lambda. Essentially, if we see firms are waiting longer than expected to replace their machines via the machines' ages seen in the data, we can use this information to determine λ. In particular, by using the distribution of the unobservables, ϵ_{it}, and other existing assumptions and information made and given about our original model, we can calculate the expected age of when a firm should replace its machine. Therefore, if we see across all observations/firms that engines have an age older than the expected value (which is when an engine is expected to be replaced), this will allow us to learn λ.

8).

Using pseudo-log-likelihood to estimate our parameters (θ_1, R), we obtain the following values for $(\widehat{\theta_1}, \hat{R}) =$ (-1.1495, -4.4537). Recall that our estimates using value function iteration are

(-1.1484, -4.4464). We can see that the two estimates are quite close to each other already, and only very slightly off. We did not calculate standard errors because of the difficulty in doing so when involving pseudo-likelihoods.

9).

After two iterations of the AM iteration process of the Hotz-Miller approach, we see that our parameter estimates for (θ_1, R), $(\widehat{\theta_1}, \hat{R})$, are (-1.1484, -4.4465). This is nearly identical to the parameter estimates obtained by MLE and value function iteration in problem 5. We did not calculate standard errors because of the difficulty in doing so when involving pseudo-likelihoods.

10).

The key difference between this problem and problem 9 is that our initial value for conditional probabilities $\widehat{p(\cdot)}$ are not from observed data, but instead chosen randomly. Specifically, we drew probabilities from a standard uniform distribution. Despite this difference, we find that still after two iterations, our parameter estimates for (θ_1, R), $(\widehat{\theta_1}, \hat{R})$, are (-1.1484, -4.4464). This is nearly identical to the parameter estimates obtained by MLE and value function iteration in problem 5. We did not calculate standard errors because of the difficulty in doing so when involving pseudo-likelihoods.