Structural Estimation in Applied Microeconomics
Problem Set 2

This problem set is designed to give you practical knowledge of the economic models and econometric methods discussed in our lectures on single agent dynamics. You will have the option to do a referee report or answer to the questions in Section 2.

1 Referee Report

Select one of the papers from the syllabus section on single agent dynamic models, excluding the ones explained in class (Rust (1987) and Keane & Wolpin (1997)). Write a referee report of at most 1,500 words.

2 Single Agent Finite Horizon Dynamic Model

This exercise is based on a simplified version of Keane & Wolpin (1997). Download from the course website the file "PS2_Data.dta". This is a panel with 500 individuals followed for 10 years, who can choose every year between going to school, working as white collar or working as blue collar. They all start with 10 years of schooling and can reach at most 15, and they all start with 0 years of experience in both types of jobs. There are 9 variables in the dataset: indiv (individual ID), age, schooling decision (SchD), white collar work decision (WhCoD), blue collar work decision (BlCoD), school experience (SchE), white collar work experience (WhCoE), blue collar work experience (BlCoE).

1. Open the dataset "PS2_Data.dta" in Stata to do some descriptive analysis. Replicate Table 1 (choice distribution) and 2 (transition matrix) from Keane & Wolpin (1997). Reproduce a line graph similar to Figure 1, with a line for each career option. Looking at the descriptives you produced, what can you say about the relationship between age and choices? What about persistence and state dependance?

2. Assume individuals make decisions based on a standard human capital model, where each career option gives them a period-specific reward. How would you structure these choice-specific reward functions based on the choice distribution and transition matrix you observe in the data?

3. Load the data "PS2_Data.csv" in Matlab. The order of the variables is the same as in the Stata dataset. Let the reward functions for each individual be the following:

$$\begin{aligned} R_s(a) &= \beta_0 + \beta_1 \mathbf{I}[g(a) \geq 12] + \varepsilon_s(a) \\ R_w(a) &= \alpha_1 g(a) + \alpha_2 x_w(a) + \alpha_3 x_w^2(a) + \varepsilon_w(a) \\ R_b(a) &= \eta_1 g(a) + \eta_2 x_b(a) + \eta_3 x_b^2(a) + \varepsilon_b(a) \end{aligned} \tag{1}$$

where $g(a)$ is the school attainment at age a, $x_w(a), x_b(a)$ are white and blue collar experience, and $\varepsilon_k(a)$ are choice $k = \{s, w, b\}$ specific shocks distributed as type 1 extreme value. Let the

observed state space be $S(a) = [g(a), x_w(a), x_b(a)]$. Make a random guess of the parameters and find the alternative specific value functions for each possible path of choices and age, defined as:

$$\begin{array}{rcl} V_k(S(a), a) & = & R_k(S(a), a) + \delta E[V(S(a+1), a+1)|S(a), d_k(a) = 1], \text{ for } a < A, \\ V_k(S(A), A) & = & R_k(S(A), A), \end{array} \tag{2}$$

where A is the last year of age in the sample and d_k is the alternative-specific decision.

4. Use the full solution of the dynamic programming problem you just derived (inner loop) to estimate the parameters of the reward functions by maximum likelihood (outer loop) (note: no need for simulation here due to the distributional assumption of the shocks). Don't estimate the discount factor, just set it to $\delta = 0.95$. Calculate the standard errors using finite difference approximation of the likelihood function at the estimated parameters (hint: you can use the same code provided for the infinite horizon single agent dynamic problem explained in class). Report your results. How do you interpret these coefficients? How does this compare to your interpretation of the descriptives in questions 1 and 2?

5. Estimate the model now with unobserved heterogeneity. Assume there are 2 types of individuals in the sample, who differ in the effect of schooling on rewards from working. One type has a higher return from schooling in a white collar job and a lower return in a blue collar one, as she specialized in managerial studies, whereas the opposite is true for the other type, who specialized in a more blue collar oriented education. Don't estimate type proportions but just set them equal to 0.5. Report estimated coefficients and standard errors. How do your results compare with the previous question? How much does unobserved heterogeneity matter?

6. Produce 3 graphs similar to the one you did in question 1, one for each career decision. In these graphs plot a line for the actual data, and two the model's prediction both with and without unobserved heterogeneity. Use the predicted probabilities to construct the lines for the two models. Which model seems to fit the data better?

References

Keane, M. & Wolpin, K. (1997), 'The career decisions of young men', *Review of Economics and Statistics* **105**(3), 473–522.

Rust, J. (1987), 'Optimal replacement of gmc bus engines: An empirical model of harold zurcher', *Econometrica* **55**, 999–1033.