Econ 8602: Wages and Employment, Applied Job Search Models
University of Minnesota

Problem set 1:
Maximum Likelihood Estimation (MLE) of the Burdett and Mortensen (1998) model

Introduction and description of the data. The main aim of this exercise is to estimate a simple version of the Burdett and Mortensen (1998, hereafter BM) steady-state wage posting model, and to obtain predictions from that model. The estimation protocol is borrowed from Bontemps, Robin and Van den Berg (1999).

The file `BM_data.dta` (in Stata 13 format) or `BM_data.csv` (in comma separated format) contain information about a sample of workers from the British Household Panel Survey (BHPS). The sample comprises an initial cross-section of 2,263 British workers present in the BHPS data in the year 1992, aged between 25 and 60 at that point. These workers are followed from the date of their 1992 interview (interviews occur throughout the year), up to their first observed labor market transition (or until they are last observed in the sample if no transition is recorded — the final sample year is 2003 and there is attrition, assumed to be random, before that final year).

Workers are initially either employed or unemployed (workers observed in self-employment, retirement, family care, sick leave, etc. are dropped from the sample). Employed workers can experience two types of transition: job-to-unemployment (J2U) or job-to-job (J2J). Unemployed workers can only transit from unemployment to a job (U2J).

For each worker in the sample, information is recorded about at least one, and at most two labor market spells, $k = 1$ or 2. The following information is available in the data file `BM_data.dta`:

variable name	type	description
`pid`	long	cross-wave person identifier
`year`	byte	year
`jbsect`k	byte	employment sector in spell $k = 1, 2$
`jbhrs`k	byte	working hours in spell $k = 1, 2$
`age`	int	age
`e`k	byte	employment status in spell $k = 1, 2$ (employed coded as 1)
`spelldur1`	double	duration of spell 1 (months)
`u2j`	byte	indicator of U2J transition
`j2j`	byte	indicator of J2J transition
`j2u`	byte	indicator of J2U transition
`sex`	byte	sex (male coded as 1)
`educ`	byte	education (highest level)
`nonwhite`	byte	race (non-white coded as 1)
`logw1`k	double	log wage in spell $k = 1, 2$ (CPI-deflated)

Missing values (coded as a "dot" in Stata) occur for all spell-2 variables when only one spell is recorded (i.e. when the worker does not experience a labor market transition during the observation window). The variable `educ` is coded as follows: 1 = less than A-levels, 2 = A-levels, 3 = some higher education. The variables `jbsect`k, and `jbhrs`k are original variables from the BHPS — see `https://www.iser.essex.ac.uk/bhps/documentation` for information.

SECTION I. Preliminary data analysis.

1. Visit the BHPS website (`https://www.iser.essex.ac.uk/bhps`) and familiarize yourself with the basic structure and contents of the BHPS data. What features make it a suitable data set for the estimation of the BM model?

2. Open the file `BM_data.dta` and create a do-file in which you will code answers to the following questions (please call it something like `ECON34610_assignment_`*your_name*`.do`).

   (a) What is the sample size? What is the sex ratio in the sample?

   (b) What is the sample unemployment rate? What is the sample unemployment rate of men? Of women? Or workers in each education category?

   (c) What proportion of initial spells are right-censored? Answer the same question for each type of first spell (job or unemployment spell).

3. Still in the same do-file, construct the initial (spell-1) cross-sectional CDF (call it `G`) and density (call it `Gdens`) of log wages `logw1`. Produce plots of those two objects.

4. Create a variable categorizing `logw1` into 25 bins (ie, percentiles 1–4, 5–8, 9–12, ..., 97–100), and a variable containing the mean spell-1 duration (`spelldur1`) within each of these 25

bins. Plot those mean durations against the wage percentiles. Is this consistent with the BM model?

5. Explain how one can obtain a non-parametric estimate of the wage sampling distribution F from the data in `BM_data.dta`. Construct this non-parametric estimate, and plot it on the same graph as `G`. Is this consistent with the theory? What else can you say about this estimate of F?

SECTION II. Estimation.

1. Open, read, and run the do-file `export_to_matlab.do`. This will export the variables `e`, `logw1`, `G`, `j2j`, `j2u`, `u2j`, and `spelldur1` in that specific order from `BM_data.dta` into an ascii file that can be read by another compute package, in this case MATLAB.

2. Write MATLAB code for the MLE estimation of the BM model following the two-step protocol of Bontemps, Robin and Van den Berg (1999). Please put all your m-files in a folder called `ECON34610_assignment_matlab_`*your_name*.

3. Write MATLAB code computing the standard errors of the estimates of δ, λ_0 and λ_1, explaining the assumptions upon which those standard errors rely.

4. The file `BM_data_simulated.mat` or `BM_data_simulated.csv` contains artificial data resulting from a simulation of 5,000 workers behaving according to the BM model with parameters $\delta = 0.01$, $\lambda_0 = 0.1$, $\lambda_1 = 0.05$ (monthly values). The simulated data has the same format as the "real-world" data used in previous questions: it comes as a structured array called `dat` with fields `dat.e`, `dat.logw1`, `dat.G`, `dat.j2j`, `dat.j2u`, `dat.u2j`, and `dat.spelldur1`. Run your ML estimation routine on the simulated data, and check that your estimates against the true parameter values.

SECTION III. Playing around with the model.

1. What is the predicted unemployment rate from the estimates obtained in Section II? Compare it with the sample unemployment rate, and discuss the possible reasons for any discrepancy.

2. Construct kernel density estimates of the cross-section distribution of wages $g(w)$ and of the sampling distribution $f(w)$. Plot both densities on the same graph.

3. Construct the distribution of firm productivity that rationalizes the observed wage distribution within the BM model. Plot firm productivity against wages, and against the cross-section CDF of wages $G(w)$. Do you notice anything wrong?

4. Looking at the predicted profit rate of high-productivity firms, what else can you say about the BM model?