Homework 1: Labor Economics

Dhananjay Ghei

December 3, 2018

1 Preliminary analysis

1. Visit the BHPS website and familiarise 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 and answer the following questions:

   (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).

We have a sample of 2263 workers, 54.18% of which are male and 45.82% are female. The sample unemployment rate is 5.92%. Table 1 shows the unemployment rate on the basis of sex. The unemployment rate is higher in males compared to females. Table 2 shows the unemployment rate on the basis of education. The unemployment rate is the highest in people with less than A-level education and lowest in people with some higher education.

Table 1 Unemployment rate (by sex)

The table shows the unemployment rate in the BHPS sample on the basis of sex.

Sex	Unemployed	Total	Unemployment Rate
Female	39	1037	3.76
Male	95	1226	7.75

Table 2 Unemployment rate (by education)

The table shows the unemployment rate in the BHPS sample on the basis of education.

Education	Unemployed	Total	Unemployment Rate
less than A-levels	97	1453	4.29
A-levels	22	361	0.97
Some Higher Education	15	449	0.66

3. Construct the initial (spell-1) cross-sectional CDF (`G`) and density of log wages `logw1`. Produce the plots of these two objects.

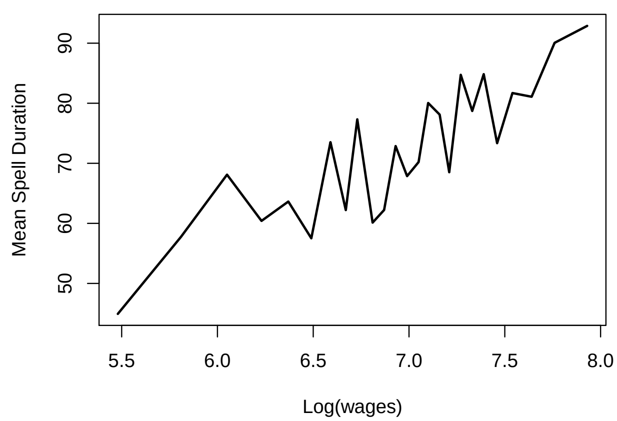
Figure 2 shows the empirical CDF of G estimated using the data on log wages of workers who were employed in the initial spell. Figure 3 shows the kernel density estimates of g and f using the same data as above. The kernel density estimate is a non-parametric estimator of the density function.

4. Create a variable categorising `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?

Figure 1 shows the mean durations against the wage percentiles. It is evident from the figure that the duration has an increasing trend along the wage percentiles. Therefore, one would posit that as wages increase, the duration in the job also increases as it takes more time for firms to compensate the worker with the same wage offer as their current ones. Table 3 in Appendix A gives the data used for generating the plot below.

Figure 1 Variation in mean duration across wage percentiles

The figure shows the mean durations against the wage percentiles. The solid line shows the actual mean wage duration in the wage percentile.



5. Explain how one can obtain a non-parametric estimate of the wage sampling distribution F from the data. 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 the estimate of F?

As mentioned in the slides, the cdf of wages accepted by workers who were just hired from unemployment provides a direct estimator $\hat{F}$ of the sampling distribution. I construct the empirical CDF of F using the wages of newly hired from unemployment. Figure 2 shows the empirical CDF of F and G estimated non-parametrically using the data on log wages. The solid gray line shows the empirical distribution function of $\hat{F}$ and the solid black line shows the empirical distribution function of $\hat{G}$. This is consistent with the theory that we learnt in class. Figure 2 shows that $\hat{G}$ first order stochastically dominates $\hat{F}$. This is expected given the selection of workers into better jobs.

2 Estimation

1. Write code for the MLE estimation of the BM model following the two-step protocol of Bontemps, Robin and Van den Berg (1999).

Figure 2 Non-parametric estimate of empirical CDF ($\hat{F}$ and $\hat{G}$)

The figure shows the empirical cdf of F and G estimated from the data on wages using BHPS survey data. The solid gray line shows the empirical distribution function of $\hat{F}$ and the solid black line shows the empirical distribution function of $\hat{G}$.

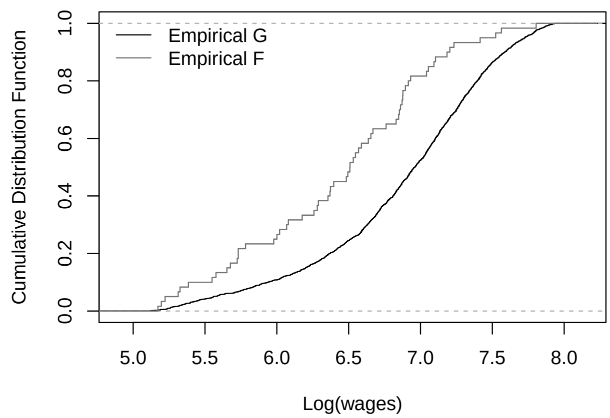
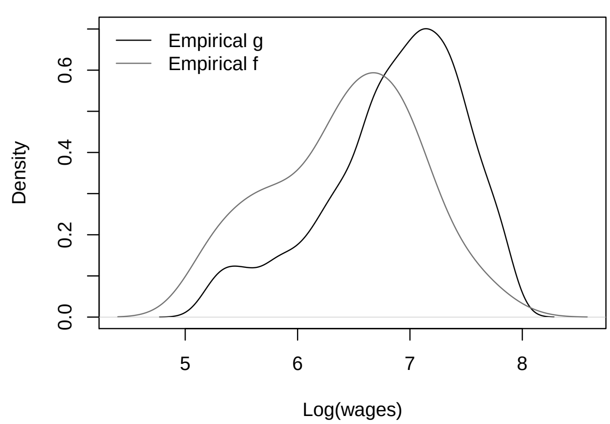


Figure 3 Kernel density estimate of PDF ($\hat{f}$ and $\hat{g}$)

The figure shows the empirical pdf of f and g estimated from the data on wages using BHPS survey data. The solid gray line shows the kernel density estimate of $\hat{f}$ and the solid black line shows the kernel density estimate of $\hat{g}$.



I follow the two-step protocol as laid out in Bontemps, Robin and Van den Berg (2000)[1]. Given our estimates of $\hat{G}$ and $\hat{g}$, I write functions $\hat{F}(:,\kappa_1)$ and $\hat{f}(:,\kappa_1)$ as functions of $\kappa_1 = \frac{\lambda_1}{\delta}$. These expressions are given as:

$$\hat{F}(w) = \frac{(1+\kappa_1)\hat{G}(w)}{1+\kappa_1\hat{G}(w)} \qquad \text{and} \qquad \hat{f}(w) = \frac{(1+\kappa_1)\hat{g}(w)}{[1+\kappa_1\hat{G}(w)]^2}$$

Next, I write the likelihood function for the employed and unemployed using the expressions below:

$L(x_i|e_1 = 0) = \lambda_0^{1-c_i} exp(-\lambda_0 d_i) f(y_{0i})^{1-c_i}$

$L(x_i|e_i = 1) = g(y_{i1})(\delta + \lambda_1 \bar{F}(y_{i1}))^{1-c_i} exp((-\delta + \lambda_1 \bar{F}(y_{i1}))d_i) \left(\frac{\delta}{\delta + \lambda_1 \bar{F}(y_{i1}}\right)^{\tau_{JUi}} \left(\frac{\lambda_1 \bar{F}(y_{i1})}{\delta + \lambda_1 \bar{F}(y_{i1})}\right)^{\tau_{JJi}}$

The generic likelihood contribution of an observation x_i is therefore:

$$L(x_i) = \left(\frac{\lambda_0}{\delta+\lambda_0} L(x_i|e_i = 1)\right)^{e_i} \left(\frac{\delta}{\delta+\lambda_0} L(x_i|e_i = 0)\right)^{1-e_i}$$

which is a function of F, G and the parameters δ, λ_0, and λ_1 Finally, I take the natural logarithms of these, add them up and minimise the negative of the log likelihood using a routine optimisation in `R`. The solver converges and the values for these parameters are given by: $\lambda_0 = 0.1496$, $\lambda_1 = 0.01359$, and $\delta = 0.01$ which gives a value of $\kappa_1 = \frac{\lambda_1}{\delta} = 1.3591$.

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

---

3. The file `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). Run your ML estimation routine on the simulated data, and check that your estimates against the true parameter values.

Note that the simulated data does not have a column for wages of unemployed who were hired. But, from the lecture slides, one can see that if wage is unobserved, we can integrate out the wages from the likelihood expression and re-write it as:

I use this fact and then run the ML estimation routine as before on the simulated data set. My parameter estimates are $\lambda_0 = 0.09986$, $\lambda_1 = 0.05385$ and $\delta = 0.00999$. These are close to the true parameter values. Therefore, it gives us a sanity check as to if the ML estimation routine works well enough or not.

3 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.

Given our estimates of the parameters and the steady state assumption, the sample unemployment rate can be calculated as:

$$Pr\{e_i = 0\} = 1 - Pr\{e_i = 1\} = \text{Unemp. rate} = \frac{\delta}{\delta + \lambda_0}$$

[1] I think the paper you wanted to refer to was the 2000 one and not the 1999 one. The lecture slides say that it is the 2000 one.

For BHPS data, the sample unemployment rate is 5.92% whereas the predicted unemployment rate from the MLE estimation of parameters is 6.26%. For simulated data, the sample unemployment rate is 9.58% whereas the predicted unemployment rate from the MLE estimation is 9.09%.

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.

Figure 3 shows the kernel density estimates for the BHPS data. These kernel density estimates are from the sample itself and not from the predictions of our estimation routine.

3. Construct the distribution of firm productivity that rationalises 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?

A Mean duration across wage percentiles

Table 3 Variation in mean duration across wage percentiles

Quartiles	Avg. Spell Duration
Q1	44.930
Q2	57.770
Q3	68.109
Q4	60.413
Q5	63.635
Q6	57.518
Q7	73.517
Q8	62.215
Q9	77.323
Q10	60.128
Q11	62.230
Q12	72.871
Q13	67.869
Q14	70.206
Q15	80.044
Q16	78.110
Q17	68.517
Q18	84.736
Q19	78.715
Q20	84.845
Q21	73.353
Q22	81.702
Q23	81.081
Q24	90.059
Q25	92.881