Empirical Methods for Policy Evaluation
Homework

Bin Cheng

December 2021

Job search with formal and informal jobs

We denote the value function for an unemployed individual as:

$$\rho U = b + \lambda_f \int \max\left[E_f(w), U\right] dG_f(w) + \lambda_i \int \max\left[E_i(w), U\right] dG_i(w) - (\lambda_f + \lambda_i)\, U \tag{1}$$

where the present discounted value function of an employed individual in sector $j, \in \{i, f\}$ given wage ω is

$$E_j(w) = \frac{w_j + \eta_j U}{\rho + \eta_j}, \quad j \in \{i, f\} \tag{2}$$

Question 1

To derive the analytical expression for the reservation wage, we define reservation wage as $\omega_j^*, j \in \{i, f\}$, which is pinned down by

$$\omega_j^* = \rho U, j \in \{i, f\} \tag{3}$$

We also define two superlevel sets L^+_{j,w^*} for each sector as

$$L^+_{j,w^*}(w) = \{w_j : w_j > w^*\}, \quad j \in \{i, f\} \tag{4}$$

Then we substitute (2), (3) and (4) into (1) and obtain

$$w_j^* = b + \frac{\lambda_f}{\rho + \eta_f} \int_{L^+_{f,w^*}(w)} \left(w_f - w_j^*\right) dG_f(w) + \frac{\lambda_i}{\rho + \eta_i} \int_{L^+_{i,w^*}(w)} \left(w_i - w_j^*\right) dG_i(w), j \in \{i, f\} \tag{5}$$

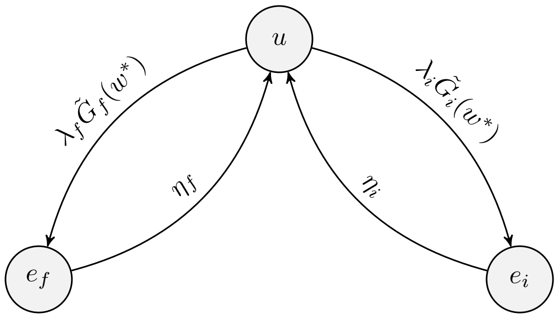
Easily we can find that reservation wage $\omega_j^*, j \in \{i, f\}$ is the same for employment in both sectors by monotonicity. This originates from the fact that once individuals were offered wages in any sector, their decisions about employment and unemployment only depend on wage in that sector, which drives the reservation wage to be the same across sectors.

This reduce (5) to

$$w^* = b + \sum_{j \in \{i,f\}} \frac{\lambda_j}{\rho + \eta_j} \int_{w^*}^{\infty} (w - w^*)\, dG_j(w) \tag{6}$$

Question 2

The job search dynamics can be viewed as a continuous-time Markov chain with three states. We present the diagram as follows



where $\tilde{G}_j(w^*)) = 1 - G_j(w^*), j \in \{i, f\}$.

To derive the invariant distribution $\mu = (p(u), p(e_i), p(e_f))^T$ of each state $s \in S = \{u, e_i, e_f\}$, we define the unique corresponding transition rate matrix as

$$Q = \begin{pmatrix} -(\lambda_i\tilde{G}_i(w^*) + \lambda_f\tilde{G}_f(w^*)) & \lambda_i\tilde{G}_i(w^*) & \lambda_f\tilde{G}_f(w^*) \\ \eta_i & -\eta_i & 0 \\ \eta_f & 0 & -\eta_f. \end{pmatrix} \tag{7}$$

Apparently Q is irreducible and recurrent. By Theorem 3.5.5 in Norris (1997), we have

$$\mu^T Q = 0 \tag{8}$$

where $\mu^T(1, 1, 1) = 1$.

Thus, it reduces to

$$\begin{aligned} p(u) &= \frac{1}{1 + \lambda_f\tilde{G}_f\eta_f^{-1} + \lambda_i\tilde{G}_i\eta_i^{-1}} \\ p(e_j) &= \frac{\lambda_j\tilde{G}_j\eta_{-j}^{-1}}{1 + \lambda_f\tilde{G}_f\eta_f^{-1} + \lambda_i\tilde{G}_i\eta_i^{-1}}, j \in \{i, f\} \end{aligned}$$

Interestingly, equation (8) shows that

$$\lambda_j\tilde{G}_j p(u) = \eta_i p(e_j) \tag{9}$$

which means that in steady state equilibrium, the ratio(number) of individuals exiting unemployment into any sector, because they get wage offer above reservation wage from that sector, must equal the ratio(number) of individuals entering unemployment from that same sector.

Question 3

An observation in the market can be either unemployed with duration $\tilde{t}_u$ or employed in informal or formal sector with wage ω. From (7), the hazard rate of exiting from unemployment is $\sum_{j\in\{i,f\}} \lambda_j\tilde{G}_j$.

Thus, we write down the log likelihood function for individuals in different states as follows

$$L(u, \tilde{t}_u) = p(u) f_u(\tilde{t}_u) = \frac{\left(\lambda_f\tilde{G}_f + \lambda_i\tilde{G}_i\right)\exp\left(-\left(\lambda_f\tilde{G}_f + \lambda_i\tilde{G}_i\right)\tilde{t}_u\right)}{1 + \lambda_f\tilde{G}_f\eta_f^{-1} + \lambda_i\tilde{G}_i\eta_i^{-1}} \tag{10}$$

$$L(e_j, w) = p(e_j)\frac{g_j(w)}{\tilde{G}_j(w^*)} = \frac{g_j(w)\lambda_j\eta_j^{-1}}{1 + \lambda_f\tilde{G}_f\eta_f^{-1} + \lambda_i\tilde{G}_i\eta_i^{-1}}, \quad j \in \{i, f\} \tag{11}$$

Thus a sample with $\left\{\tilde{t}_u(l)\right\}_{l=1}^{N_u}$ and $\{w_j(l)\}_{l=1}^{N_{e_j}}, j \in \{i, f\}$ have log likelihood function as

$$
\begin{aligned}
\ln L &= \sum_{l \le N_u} \ln\left(L\left(u, \tilde{t}_u(l)\right)\right) + \sum_{j\in\{i,f\}} \sum_{l\le N_{e_j}} \ln\left(L\left(e_j, w_j(l)\right)\right) \\
&= -(N_u + N_{e_i} + N_{e_f}) \ln\left(1 + \lambda_f G_f\left(w^*\right)\eta_f^{-1} + \lambda_i \tilde{G}_i\left(w^*\right)\eta_i^{-1}\right) \\
&+ \sum_{j\in\{i,f\}} N_{e_j} \ln\left(\lambda_j \eta_j^{-1}\right) + N_u \ln\left(\sum_{j\in\{i,f\}} \lambda_j \tilde{G}_j\left(w^*\right)\right) \\
&- \left(\sum_{j\in\{i,f\}} \lambda_j \tilde{G}_j\left(w^*\right)\right) \sum_{l\le N_u} t_u(l) + \sum_{j\in\{i,f\}} \sum_{l\le N_{e_j}} \ln\left(g_j\left(w_j(l)\right)\right)
\end{aligned} \tag{12}
$$

Question 4

By symmetry we can have the following system of log likelihood functions' first-order conditions with respect to job offer and termination rates $(\lambda_j, \eta_j), j \in \{i, f\}$ as follows*

$$
\frac{\partial \ln L}{\partial \lambda_j} = -N \frac{\tilde{G}_j\left(\omega^*\right)\eta_j^{-1}}{1 + \lambda_f \tilde{G}_f\left(\omega^*\right)\eta_f^{-1} + \lambda_i \tilde{G}_i\left(\omega^*\right)\eta_i^{-1}} + \frac{N_{e_j}}{\lambda_j} + \frac{N_u \tilde{G}_j\left(\omega^*\right)}{\sum_{k\in\{i,f\}} \lambda_k \tilde{G}_k\left(w^*\right)} - \tilde{G}_j\left(\omega^*\right) \sum_{l\le N_u} t_u(l) = 0 \tag{13}
$$

$$
\frac{\partial \ln L}{\partial \eta_j} = \frac{N}{\eta_j^2} \frac{\tilde{G}_j\left(\hat{w}^*\right)\lambda_j}{1 + \lambda_f \tilde{G}_f\left(\hat{w}^*\right)\eta_f^{-1} + \lambda_i \tilde{G}_i\left(\hat{w}^*\right)\eta_i^{-1}} - \frac{N_{e_j}}{\eta_j} = 0 \tag{14}
$$

where $j \in \{i, f\}$.

We substitute (14) into (13) and find that

$$
\frac{N_u}{\sum_{k\in\{i,f\}} \lambda_k \tilde{G}_k\left(w^*\right)} = \sum_{l\le N_u} t_u(l) \tag{15}
$$

which is independent from j.

Thus we only have 3 valid first-order conditions for identification, instead of 4. And we can only identify 3 parameters without further assumptions.

Question 5

We assume homogeneous job offer and termination rate λ, η and log-normal distribution of wage offer. Then (14) and (15) are simplified into

$$
\lambda = \frac{N_u}{\sum_{j\in u} t_u(j)} \frac{1}{\sum_{k\in\{i,f\}} \tilde{G}_k\left(w^*\right).} \tag{16}
$$

$$
\eta = \lambda\left(N \frac{\tilde{G}_j}{N_{e_j}} - \tilde{G}_f - \tilde{G}_i\right) \quad j \in \{i, f\} \tag{17}
$$

In addition, (17) also leads to the equality (18).

$$
\frac{\tilde{G}_i}{N_{e_i}} = \frac{\tilde{G}_f}{N_{e_f}} \tag{18}
$$

Later, we will test optimality of MLE by comparing two sides of (16) with estimated parameters as (16) contain the most information about unemployment duration time, wage and distributional parameters.

*We denote $N = N_u + N_{e_i} + N_{e_f}$ as the total stock of market.

Step 1: Identify Reservation Wage ω^*

We propose an estimator of ω^* as $\hat{\omega}^* = \min\{\{w_j(l)\}_{l=1}^{N_{e_j}}, j \in \{i, f\}\}$ given homogeneity of reservation wage in two sectors $j \in \{i, f\}$ as (6).

Step 2: Inspect Population Figure

	Chile	Argentina	Colombia	Mexico
N_u	873	400	607	458
%	11.57	8.39	16.40	7.08
N_{e_i}	865	1773	1311	3641
%	11.46	37.19	35.41	56.28
N_{e_f}	5807	2594	1784	2371
%	76.96	54.42	48.19	36.65
N	7545	4767	3702	6470

Table 1: Employment and Unemployment Population Figures

We provide population statistics about each country as summarized in Table 1. And we point out several interesting facts helping understand MLE estimates better

- Mexico has the lowest unemployment rate 7.08% while Colombia has the largest unemployment rate at 16.4%.
- Mexico obtains the highest informal employment rate at 56.28% and the lowest formal employment rate at 36.65%. Quite on the contrary, Chile owns the lowest informal employment rate at 11.46% and the highest formal employment rate at 76.95%.

Step 3: MLE

	Chile	Argentina	Colombia	Mexico
μ_i	-3.5514	0.7720	0.0106	0.1117
	(0.1207)***	(0.0126)***	(0.0107)	(0.0073)***
σ_i	2.6692	0.5295	0.3856	0.4392
	(0.0817)***	(0.0091)***	(0.0075)***	(0.0053)***
μ_f	0.9146	1.3995	0.2259	0.2874
	(0.0047)***	(0.0089)***	(0.0065)***	(0.0087)***
σ_f	0.3604	0.4537	0.2765	0.4233
	(0.0033)***	(0.0063)***	(0.0046)***	(0.0062)***
λ	0.3381	0.1801	0.1591	0.3455
	(0.0116)***	(0.0090)***	(0.0065)***	(0.0162)***
η	0.0512	0.0330	0.0624	0.0526
	(0.0025)***	(0.0024)***	(0.0038)***	(0.0035)***
b	-24.6605	-13.1817	-2.5688	-6.8501
ω^*	0.4169	0.3762	0.2119	0.2855
$E(\omega_i)$	1.0110	2.4897	1.0886	1.2314
$SD(\omega_i)$	35.6194	1.4162	0.4358	0.5680
$E(\omega_f)$	2.6634	4.4923	1.3023	1.4579
$SD(\omega_f)$	0.9920	2.1476	0.3670	0.6458

Table 2: Parameters Estimated by MLE with SE in Parenthesis. $*, p < 0.1$; $**, p < 0.05$; $***, p < 0.1$

To capture global maximum of log likelihood function, we uniformly draw our initial guess of parameter set $\left((\mu_j, \sigma_j)_{j\in\{i,f\}}, \lambda, \eta\right)$ from a hyper cubic $\prod_{k=1}^{6}[a_k, b_k]$.

Then we employ several different MLE methods like *optim(),maxLik()* and discover that they deliver extremely similar results. For simplicity, we only present results with package *maxLik()*. In workspace image

Result Collection.RData, matrices *LLK_maxLik_country* stored parameter values and corresponding loglikelihood, which shows robustness of convergence to global maximum for such a large set of distinct and randomly drawn initial guesses.

Step 4: Interpretation

The statistics reported in Table 2 are consistent with our population figures.

To understand the job offer and termination rate as a pair (λ, η), we find that, endowed with the highest unemployment rate among four countries, Colombia has also the lowest job offer rate 0.1591 and the highest job termination rate 0.0624. Thus, these statistics naturally support the fact that the faster job being offered than job being terminated, the lower unemployment rate tend to be. Country like Mexico can serve as a perfect example as that it has the largest difference $\lambda - \eta$ among four countries, and endowed with the lowest unemployment rate.

Next, the relationship between reservation wage ω^* and benefit from unemployment b is very simple in that the higher reservation wage is, the lower benefit from unemployment become. This can be viewed as a natural result of model setting. Workers are assumed to be homogeneous and search is random, which means that workers' opportunity cost, when refuse a current job offer for higher wage in the future, enlarge with higher bar of reservation wage. As for the sign of b, it can result from the fact that higher wage job offers in the future fail to exist for all seekers, which leave the benefit from unemployment mostly cost, thus negative instead of positive. Then since a worker is more likely to be satisfied with and thus take a current job offer in Colombia than in Chile, given that the reservation wage is much lower, it helps explain the highest b in Colombia (lowest opportunity cost).

Then we focus on large cross-country difference in wage returns to formality and informality. An immediate observation is that Argentina has the highest expected wage in both informal and formal sector among countries. And formal sector always has a higher expected wage than informal sector for all countries and naturally higher expectation of value of employment as well.

In addition, the difference between expected wage and value is also consistent with the difference between employment rate in two sectors. It turns out that the more employment rate in formal sector than in informal sector, the higher the difference between expected wages and values. In other words, more people choose to work in formal sector when such choice bring higher value and wage. For example, more than 76% people in Chile are employed in formal sector while only 11% people are in informal sector. This result from the fact that expected value given expected wage in formal sector is 30.5, much higher than that in informal sector 14.2 as shown in Table 3.

	Chile	Argentina	Colombia	Mexico
$E_{e_i}(E(\omega_i))$	14.2076	32.9942	12.0382	14.9283
$E_{e_f}(E(\omega_f))$	30.5315	57.1290	13.9393	17.1349

Table 3: Expected Values given Expected Wage in Sectors

Step 5: Optimality Test

	Chile	Argentina	Colombia	Mexico
λ	0.3381	0.1801	0.1591	0.3455

Table 4: λ generated by first-order condition (16)

Lastly, we test optimality of our estimates. And we substitute estimed lognormal distribution parameters $\mu_j, \sigma_j, j \in \{i, f\}$, as shown in 1-4th rows of Table 2, into the right hand side of (16). It turns out that, by comparing row 5th of Table 2 and Table 4, the values of λ induced by first order condition fit the values of λ estimated by MLE pretty well.

Thus, the validity of MLE estimates holds.