Empirical Methods for Policy Evaluation

TSE PhD Program

Takehome 2

Directions. You can work in pairs, although such pairs have to differ from those of Takehome 1 (Augustin's part). Send your answers in a pdf file along with the code you used to produce the estimates to matteo.bobba@tse-fr.eu by **January 14, 2022**. Do not submit your work after the deadline.

Job search with formal and informal jobs

Consider the following simple variant of the partial-equilibrium job search model that we saw in class. Jobs can be formal (f) or informal (i). Poisson rates of job offers and job termination are λ_j and η_j, respectively, and wages are drawn from sector-specific exogenous stationary distributions $G_j(w)$, with $j \in \{i, f\}$. The value function for an unemployed individual is given by:

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

where ρ is the discount factor, b is the flow value of being unemployed, and $E_j(w) = \frac{w_j + \eta_j U}{\rho + \eta_j}$ is the present discounted value of employment in sector $j \in \{i, f\}$.

Question 1: Derive the analytical expression for the reservation wage in this economy. Does it vary between formal and informal jobs? Why?

Question 2: Derive the steady-state proportions of formal employees (e_f), informal employees (e_i) and unemployed individuals (u) in this economy.

Question 3: Write down the log likelihood function for a cross-sectional sample of individuals with (right-censored) unemployment durations[1] $\{\tilde{t}_u(i)\}_{i=1}^{N_u}$ and hourly wages in the two sectors $\{w_j(i)\}_{i=1}^{N_{e_j}}$, $j \in \{i, f\}$.

Question 4: Inspect the systems of equations obtained by the first-order conditions of the log likelihood function with respect to the hazard rates out of unemployment and the job termination rates. How many parameters can be identified in this model?

[1]Spells that are on-going at a point in time have not yet concluded, so the $t_u \geq \tilde{t}_u$ (right-censored spells). Spells that are sampled at a random point in time are likely to be longer that those drawn from the population distribution (length-biased sampling). If the population density of unemployment spells is negative exponential, then the distribution of right-censored, length-biased unemployment spells is the same as the population density of these spells, or $\tilde{f}_u(x) = f_u(x), \forall x \geq 0$.

Question 5: Estimate the parameters of the model by Maximum Likelihood using the different datasets made available in Moodle. Assume that $G_j(w) \sim \log N(\mu_j, \sigma_j^2)$, $\lambda_f = \lambda_i = \lambda$ and $\eta_f = \eta_i = \eta$. Fix ρ=0.05. The structure of the data is as follows:

- Column 1: Unemployment (ongoing) duration, defined only for unemployed individuals. The value for employees is just set to zero.
- Column 2: Hourly wages in US dollars, defined only for employed individuals. The value for unemployed is just set to zero.
- Column 3: Labor market status (1 - unemployed, 2 - formal employee, and 3 informal employee).

Report in separate tables the estimated coefficients for each of the four countries along with the associated standard errors, the implied means and standard deviations of the distributions of offered and accepted hourly wages in the two sectors,[2] and the reservation wage. Comment on your results, with special reference to the cross–country differences in the wage returns to formality.

Here are a few guidelines on coding/numerical methods:

a) Create a function that evaluates the logarithm of the likelihood using as inputs the vector of parameters and the matrix with the data. Before trying any maximization attempt, make sure that you first evaluate the log-likelihood function at initial parameters and check that it is well defined for each observation in the sample. Change the value of the initial parameters and check again.

b) All the model parameters except b are defined in $\mathbb{R}_+$. A simple trick to avoid searching over negative values is to maximize the likelihood for a transformed vector of parameters $\theta' = exp(\theta)$ and then re-express the resulting estimates as $\theta = ln(\theta')$.

c) You can use any of the maximization routine available in your software (BFGS, DFP, Newton-Raphson, BHHH, etc.). Results should not be sensitive to that choice but it's good that you check. Many optimization routines are designed for minimization. If that is the case in your software you just minimize the negative of the log-likelihood function.

d) To avoid local maximums or minimums, try with different initial values for the parameter vector and/or check that the numerical gradient changes from zero to positive or negative after perturbing the values of some of the estimated parameters.

[2]Notice that if $x \sim \log N(\mu, \sigma^2)$ then $E(x) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$ and $SD(x) = \sqrt{[\exp(\sigma^2) - 1]\exp(2\mu + \sigma^2)}$.