Final Assignment

April 9, 2020

The goal of this exercise is to study the relationship between individual observed characteristics and the value of leisure. For the rest of the exercise, we will use data from GSOEP (German Panel), from the AER library, which is a cross sectional data consisting of 675 individuals. We will be interested in the variables (kids, marital, meducation, and income). Obviously, the value of leisure is not available. As a consequence, we will develop a empirical model where wages are determined by opportunities in the labor market and leisure value. Formally, for a given value of leisure b, we will recover a reservation wage $\phi(b)$. Then, under parametric assumption on how individuals characteristics affect the value of leisure b, we will fit the wages implied under the model to the actual data. In order to accomplish this task, there is a number of pre-requisites, that are described from Exercise 1 to 4. In Exercise 5, everything comes together and the objective function is recovered.

Exercise 1 Kernel Density Estimation

Suppose you have access to a vector $X = (x_1, \dots x_n)$ (in our current example, wages), and you would like to estimate the density function without making any assumption on the distribution of X. The kernel density estimator at x_0 is given by

$$\widehat{f}_h(x_0) = \frac{1}{n}\sum_{i}^{n} K_h(x_0 - x_i) = \frac{1}{nh}\sum_{i}^{n} K\left(\frac{x_0 - x_i}{h}\right) \tag{1}$$

where n is the length of X, and h is the bandwidth.

1. Adopting silverman rule of thumb for bandwidth determination, create a function *band*, which given X, returns
$$h = \left(\frac{4\sigma^5}{3n}\right)^{\frac{1}{5}} \tag{2}$$
where σ is the standard deviation of X.

2. Adopting the gaussian kernel, create a function *gaussian*, which given X returns
$$K(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) \tag{3}$$

3. Write the function *kdens*, which given X as input, returns the kernel density estimate at point x_0.

Exercise 2 Empirical CDF

Suppose you have access to a vector $X = (x_1, \dots x_n)$, and you would like to estimate the cumulative density function without making any assumption on the distribution of X. The empirical cdf at x_0 is given by

$$\widehat{F}(x_0) = \frac{1}{n} \sum_{i}^{n} \mathbb{1}\{x_i < x_0\} \tag{4}$$

where $\mathbb{1}$ is a dummy variable.

1. Write a function *ecdf* that computes the empirical CDF at point x_0.

Exercise 3 Numerical Integration

In this exercise, we aim to approximate a solution to the following equation using trapezoidal integration rule.

$$\int_{c}^{d} f(x)dx \tag{5}$$

where $f(x)$ is the kernel density estimate and $-\infty \leq c \leq d \leq +\infty$. Recall that for a given $h = \dfrac{d-c}{n}$ and n the desired approximation grid length, the trapezoidale integration rule is given by

$$T_n(f) = h\left[\frac{f(c)}{2} + f(c+h) + f(c+2h) + \ldots + f(c+(n-1)h) + \frac{f(d)}{2}\right] \tag{6}$$

1. Write a function *trapzf* to calculate the integral using trapezoidal method.

Exercise 4 Reservation wages

In this part, we lay out a job search model to construct the reservation wage of an agent given his value of leisure b, with support $[\underline{b}, \overline{b}]$. Specifically, under a set of assumptions (continuous time, random search, ...), the value of unemployment is given by

$$rV_u(b) \quad = \quad b + \lambda_u \int_{\phi(b)}^{\overline{w}} \max\Big\{V_e(x) - V_u(b), 0\Big\} dF(x) \tag{7}$$

where λ_u is the job arrival rate for unemployed workers (the rate at which an unemployed worker meets a vacancy), $F(x)$ is the wage offer distribution (which is not known) with support $[\phi(\underline{b}), \overline{w}]$, and $\phi(b)$ is the reservation wage of a worker with leisure value b. Set $\overline{w}$ to the maximum wage in the sample.

Similarly, the value of an employed worker earning a wage w is

$$rV_e(w;b) \quad = \quad w + \lambda_e \int_{w}^{\overline{w}} \max\Big\{V_e(x) - V_e(w), 0\Big\} dF(x) + \delta\big[V_u(b) - V_e(w)\big]$$

where λ_e is the job arrival rate for employed workers (the rate at which an employed worker meets a vacancy), $F(x)$ is the wage offer distribution (which is not known), and δ is the job destruction rate. The reservation wage is such that $V_u(b) = V_e(\phi(b))$, and is explicitly given by the nonlinear equation

$$\begin{aligned}\phi(b) \quad &\equiv \quad b + \big(\lambda_u - \lambda_e\big) \int_{\phi(b)}^{\overline{w}} \overline{F}(x) dV_e(x) \\ &\equiv \quad b + \big(\lambda_u - \lambda_e\big) \int_{\phi(b)}^{\overline{w}} \frac{\overline{F}(x)}{r + \delta + \lambda_e \overline{F}(x)} dx\end{aligned}$$

Where $\overline{F} = 1 - F$. In the rest of this exercise, we will assume that $\lambda_u = 2$ (if the time unit is a year, it takes 1/2 years to find a job), $\lambda_e = 0.5$, $\delta = 0.2$ (the mean duration of jobs is 1/0.2 years), and $r = 0.01$.

Finally, the distribution of wages offered $F(x)$ is not known, but the distribution of accepted wages $G(x)$ is observed. Specifically, because workers will accept only a subset of offered jobs, these distributions will differ. However, steady state conditions imply that

$$G(w) = \frac{\delta}{\delta + \lambda_e \overline{F}(w)} \tag{8}$$

1. Create two functions *fcdf* and *fpdf* to compute the wage offer distribution using respectively the empirical cdf and kernel density estimates on the income distribution.

2. For a given b, use bisection method to solve for the reservation wage $\phi(b)$. (*Hint:* You can use pre-programmed functions for this).

Exercise 5 Indirect Inference

In this final part, we parametrize the value of leisure as

$$b = X\beta + \epsilon \tag{9}$$

Where X includes a constant, number of children (kids), marital status (marital), and education (meducation), and ϵ is an error term which follows a normal distribution $(0,\sigma)$. The set of parameters to be estimated consists β and the scale parameter σ denoted by Θ.

1. Run a regression of wages on observed attributes, and store the parameters including the sd of the

error terms. Store this outcome as $\theta_{\mathbf{data}}$

2. In this question, we construct an empirical counterpart to $\theta_{\mathbf{model}}$. To do this,

    - For a guess (initial value) $\Theta^0 = (\beta_0, \sigma_0)$, compute b_0 and recover $\phi(b_0)$. Because b includes a stochastic component, simulation methods need to be used. That is, at simulation step i, draw a vector ϵ^i, form b_0^i and recover $\phi(b_0^i)$, and

    $$\phi(b_0) = \frac{1}{S}\sum_{i=1}^{S}\phi(b_0^i) \tag{10}$$

      Where S is the number of simulations.

    - Run a regression of $\phi(b_0)$ on observed attributes, and store the parameters including the sd of the error term. Store this outcome as $\theta_{\mathbf{model}}$

3. Use numerical optimization techniques to recover the parameters Θ. Formally, we solve the following minimization problem

    $$\min_{\Theta}(\theta_{\mathbf{model}}(\Theta) - \theta_{\mathbf{data}})'(\theta_{\mathbf{model}}(\Theta) - \theta_{\mathbf{data}}) \tag{11}$$

    For identification purpose, one must ensure that $||\theta_{\mathbf{model}}|| \leq ||\theta_{\mathbf{data}}||$.

4. Recover the standard deviations by bootstrap.