Labor Field Exam

August 2019

There are two parts to this exam. Each should take about one hour. Calculators are not necessary. Please explain your notation in order to maximize chances of partial credit.

Write your answer to each part in a separate book.

Part I (244)

The work of Chetty, Friedman, and Rockoff (2014) has generated substantial interest in identifying and firing bad teachers. Suppose we have a dataset on teacher "mistakes" which might correspond to instances where a teacher fails to meet a performance goal. We suppose the number of mistakes $Y_j \geq 0$ that teacher j makes in a year can be modeled as the outcome of a series of binomial trials. That is, that:

$$Y_j \overset{iid}{\sim} Bin\left(n, p_j\right) \quad \text{for } j = 1, ..., J,$$

where p_j is teacher j's mistake probability and n is the number of performance goals teachers face in a year, which we assume is constant across teachers. Assume further that there are only two types of teachers: "good" teachers with mistake probability p_g and "bad" teachers with mistake probability $p_b > p_g$. We formalize this with the following mixture representation:

$$p_j \overset{iid}{\sim} \pi Bin\left(n, p_b\right) + \left(1 - \pi\right) Bin\left(n, p_g\right),$$

where π gives the share of teachers that are bad.

1. Derive an expression for the likelihood of observing the event $Y_j = y$.
2. Let B_j be an indicator for being a bad teacher. Derive an expression for the probability $\Pr\left(B_j = 1 | Y_j = y\right)$ that teacher j is bad given that she made y mistakes. Your expression should involve only the parameters (n, p_b, p_w, π).
3. Suppose teacher j makes y mistakes while teacher k makes $z < y$ mistakes. Given this evidence, what is the probability that teacher k has a lower probability of making mistakes than teacher j (i.e., that $p_k < p_j$)?
4. What is the probability that teacher k has a higher probability of making mistakes than teacher j (i.e., that $p_k > p_j$)? Prove that this quantity is smaller than your answer to question #3.
5. How would you estimate the parameters (p_b, p_g, π) from a dataset $\{Y_j\}_{j=1}^{J}$ of teacher mistakes? What conditions would you need to hold for this estimator to be consistent?
6. What concerns might you have with linking your estimates of this model to personnel decisions involving teachers?

PART II (250A)

Consider a labor market with heterogeneous workers and two occupations. A worker of skill θ_i earns $w_{1i} = \alpha_1 + \theta_i$ if working in occupation 1 and $w_{2i} = \alpha_2 + \gamma\theta_i$ in occupation 2. The cost of entering occupation 1 is zero, while the cost of entering occupation 2 is c. Workers choose occupations to maximize earnings net of costs.

1. Write down an equation characterizing who chooses to work in occupation 2. Do higher or lower ability individuals select into occupation 2? How does this depend on the parameters α_1, α_2, γ, and c? Provide some intuition.

2. Suppose you have data on occupation $D_i \in \{1, 2\}$ and earnings $w_i = 1\{D_i = 2\}w_{2i} + 1\{D_i = 1\}w_{1i}$ for a random sample of individuals that behave according to this model. You regress earnings on a constant and an indicator for working in occupation 2. Derive expressions for the constant and slope coefficient from this regression in terms of the parameters $\alpha_1, \alpha_2, \gamma$, and c.

3. You are interested in the return to working in occupation 2 for those who work in occupation 2, $E[w_{2i} - w_{1i}|D_i = 2]$. How does this parameter relate to the OLS slope coefficient from part (2)? Can you use theory to sign the "bias" of OLS?

4. Now suppose costs of entering occupation 2 vary across individuals. The cost of working in occupation 2 for individual i is $c_i = c_0 + \lambda\theta_i + z_i$, where z_i is independent of θ_i. Derive a new expression for the OLS slope coefficient and discuss its bias.

5. You gain access to data on the variable z_i, which takes values in {0,1}. Describe an alternative strategy to estimating the returns to working in occuption 2 using this data. Suppose this strategy indicates a higher return to occupation 2 than the OLS approach in part (4). What can you conclude?

6. A large literature in labor economics seeks to measure the causal effects of education on earnings. Discuss some patterns of findings from this literature and relate them to your answers from parts 1-5.