

Empirical Bayes Mixtape Session, Coding Lab 3: Classifying Firm-level Discrimination

This coding lab uses the data from the employment discrimination experiment of Kline, Rose and Walters (2022) to conduct an empirical Bayes multiple testing analysis. The objective of the analysis is to determine which firms can be classified as discriminating against Black applicants while controlling the False Discovery Rate (FDR).

1. Replicate steps 1-4 from coding lab 2 to obtain a data set of 108 firm-specific discrimination estimates $\hat{\theta}_f$ along with standard errors s_f .
2. For each firm, form a test statistic $z_f = \hat{\theta}_f/s_f$, and use this statistic to compute the p -value p_f from a one-tailed test of $H_0 : \theta_f = 0$ vs. $H_A : \theta_f > 0$.
3. Plot a histogram of the firm-specific p -values. What do you notice about this distribution?
4. Let $\pi_0 = \Pr[\theta_f = 0] = \int 1[\theta = 0]dG(\theta)$ denote the share of firms in the population that are not discriminating. Use your p -values to compute an upper bound $\hat{\pi}_0$ on π_0 . [Hint: compute the average height of the p -value density above some threshold λ . You can start by setting $\lambda = 0.5$, though any λ will do.] Interpret your estimated bound.
5. Use the bound from part (8) to compute q -values as $q_f = [p_f \hat{\pi}_0]/\hat{F}_p(p_f)$, where \hat{F}_p is the estimated CDF of p -values. Recall that rejecting all null hypotheses with q -values below \bar{q} controls FDR to at most $100\bar{q}\%$. Limiting FDR to 5%, how many firms can you classify as discriminating against Black applicants?