Ec143, Spring 2024

Professor Bryan Graham

Midterm Review Sheet #2

This review sheet is designed to assist you in your exam preparations. I suggest preparing written answers to each question. You may find it useful to study with your classmates. In the exam you may bring in a single 8.5 x 11 sheet of notes. No calculators or other aides will be permitted. Please bring blue books to the exam. The first midterm exam will occur in class on Thursday, March 5th.

[1] **Returns-to-schooling**

For $s \in \mathbb{S}$, a hypothetical years-of-schooling level, let an individual's potential earnings be given by $\log Y(s) = \alpha_0 + \beta_0 s + U$. Here U captures unobserved heterogeneity in labor market ability and other non-school determinants of earnings. Let the total cost of s years of schooling be given by $(A + \rho_0 U)s + \frac{\kappa_0}{2}s^2$; A is also unobserved but varies independently of U. You may assume U is mean zero, while A has a mean of δ_0.

[a] The term $A + \rho_0 U$ captures heterogeneity in the marginal costs of schooling. Explain. **[2 to 3 sentences].**

[b] Agents choose years of completed schooling to maximize expected utility

$$S = \arg\max_{s \in \mathbb{S}} \mathbb{E}\left[\log Y(s) - (A + \rho_0 U)s - \frac{\kappa_0}{2}s^2 \middle| A, U\right].$$

Show that observed schooling is given by

$$S = \gamma_0 + V, \quad \mathbb{E}[V] = -\frac{\delta_0}{\kappa_0}$$

for $\gamma_0 = \beta_0/\kappa_0$ and $V = -(A + \rho_0 U)/\kappa_0$.

[c] Show that the coefficient on schooling in the least squares fit of log earnings onto a constant and schooling will equal, when the sample size is very large,

$$b_0 = \beta_0 - \frac{\rho_0}{\kappa_0}\frac{\mathbb{V}(U)}{\left(\frac{1}{\kappa_0}\right)^2 \mathbb{V}(A) + \left(\frac{\rho_0}{\kappa_0}\right)^2 \mathbb{V}(U)}. \tag{1}$$

[d] What sign do you expect ρ_0 to take? Why? What are the implications of your assumption for the relationship between b_0 and β_0? **[3 to 4 sentences].**

[e] You consult with an economics of education expert. She says that variation in schooling levels has nothing to do with labor market ability. Interpret this statement with reference to equation (1) above **[2 to 3 sentences].**

[f] You next consult with a labor economics expert. He says that variation in schooling levels moves in lock-step with labor market ability. Interpret this statement with reference to equation (1) above **[2 to 3 sentences].**

[2] **Regression redux #1**

Let W, X be a pair of (scalar-valued) regressors with the property that $\mathbb{C}(W, X) = 0$. Show that, for outcome, Y,

$$\mathbb{E}^* [Y| 1, W, X] = \mathbb{E}^* [Y| 1, W] + \mathbb{E}^* [Y| 1, X] - \mathbb{E} [Y] .$$

You may assume that all objects in the above expression are well-defined (i.e., all necessary moments exist and so on).

[a] First show that

$$\mathbb{E}^* [\mathbb{E}^* [Y| 1, W]| 1, X] = \mathbb{E}^* [\mathbb{E}^* [Y| 1, X]| 1, W] = \mathbb{E} [Y] .$$

[b] Second verify the regression $\mathbb{E}^* [Y| 1, W, X]$ takes the claim form by checking the FOCs which define the mean squared error (MSE) minimizing linear regression. That is show that

$$\begin{aligned} \mathbb{E}[U] &= 0 \\ \mathbb{E}[UX] &= 0 \\ \mathbb{E}[UW] &= 0 \end{aligned}$$

for $U = Y - \mathbb{E}^* [Y| 1, W] - \mathbb{E}^* [Y| 1, X] + \mathbb{E} [Y]$.

[c] Finally show that

$$\mathbb{E}^* [Y| 1, W, X] = \mathbb{E} [Y] + \frac{\mathbb{C}(Y, W)}{\mathbb{V}(W)} (W - \mathbb{E} [W]) + \frac{\mathbb{C}(Y, X)}{\mathbb{V}(X)} (X - \mathbb{E} [X]) .$$

[d] The Vice Chancellor for Undergraduate Education is interested in boosting academic performance among first year students. She randomly divides first year students into two equal-sized groups. In the first group she randomly assigns half of students to receive a daily snack voucher worth $5 dollars. In the second group she randomly assigns half of students to get two hours of structured advising each semester. At the end of the semester she records

student grade point average. Explain how the Vice Chancellor can use her data to form an estimate of the best linear predictor of end-of-first year GPA given a constant, a dummy variable for snack voucher receipt *and* a dummy variable for receipt of extra advising.

[e] Under what circumstances is the linear regression computed in part [d] helpful for allocating resources across initiatives? Consider, and elaborate on, three cases: [a] snacks and advising are complements in the production of GPA, [b] they are substitutes and [c] they do not interact.

[f] Outline a more informative experiment for the Vice Chancellor. Explain why is it is "better" than the experiment described in part [d].

[3] Regression redux #2

Consider the following statistical model for the earnings of Berkeley students

$$Y = \alpha + \beta G + \gamma A + U,\ \mathbb{E}[U|G, A] = 0,$$

where G equals one if the student graduated and zero if they dropped out and A equals one if at least one of the student's parents graduated from college and zero otherwise.

[a] You read in the Oakland Tribune newspaper that Berkeley graduates earn an average of \$75,000 per year nationwide, while the earnings of dropouts average only \$15,000. Express this population earnings difference between Berkeley graduates and dropouts in terms of the statistical model given above.

[b] Under what conditions is it true that $\beta = \$60,000$? Do you think these conditions are likely to be true in practice? Briefly explain your answer [**3-5 sentences**].

[c] The same article reports that among Berkeley graduates, three fourths come from families where at least one parent completed college, while among all former students (i.e., graduates and dropouts) only seven twelfths come from such families. It also states that the overall (i.e., unconditional) graduation rate at Berkeley is two-thirds. Among dropouts, what fraction come from families where at least one parent completed college?

[d] Assume $\gamma = \$25,000$. Using your answers in parts (a) and (c) solve for β. What is the expected earnings gain associated with graduating from Berkeley holding parent's education (i.e., A) constant? Briefly comment on why your answer differs from the earnings gap between graduates and dropouts reported by the Tribune [**3-5 sentences**].

[e] You are considering dropping out of Cal to spend more time on Telegraph Avenue. What is the (approximate) expected earnings loss associated with this decision? Explain [**3-5 sentences**].

[f] You move to Oakland upon graduation, your neighbor to the left tells you that he dropped out of Berkeley during the Free Speech Movement, your neighbor to the right tells you that he graduated from Berkeley about the same time. What is your expectation of the annual earnings of your two neighbors? Explain [**3-5 sentences**].

[4] **Regression redux #3**

Let $m(Z) = \mathbb{E}[X|Z]$ and consider the linear regression

$$\mathbb{E}^*[Y|X, m(Z), A] = \alpha_0 + \beta_0 X + \gamma_0 m(Z) + A.$$

[a] Show that

$$\mathbb{E}^*[m(Z)|X] = \delta_0 + \xi_0 X$$

with

$$\begin{aligned} \delta_0 &= (1 - \xi_0)\mathbb{E}[X] \\ \xi_0 &= \frac{\mathbb{V}(\mathbb{E}[X|Z])}{\mathbb{E}[\mathbb{V}(X|Z)] + \mathbb{V}(\mathbb{E}[X|Z])}. \end{aligned}$$

[b] Assume the population under consideration is working age adults who grew up in the San Francisco Bay Area. Let Y denote a adult log income, let X denote the log income of one's parents as a child and let Z be a vector of dummy variables denoting an individual's neighborhood of residence as a child. Provide an interpretation of ξ_0 as a measure of residential stratification by income.

[c] Establish the notation $\rho = \text{corr}(A, X)$, $\mu_A = \mathbb{E}[A]$, $\mu_X = \mathbb{E}[X]$, $\sigma_A^2 = \mathbb{V}(A)$ and $\sigma_X^2 = \mathbb{V}(X)$. Show that

$$\mathbb{E}^*[Y|X] = \alpha_0 + \gamma_0(1 - \xi_0)\mu_X + \left(\mu_A - \rho\frac{\sigma_A}{\sigma_X}\mu_X\right) + \left\{\beta_0 + \gamma_0\xi_0 + \rho\frac{\sigma_A}{\sigma_X}\right\}X.$$

[d] Your research assistant computes an estimate of $\mathbb{E}^*[Y|X]$ using random sample from San Francisco. She computes a separate estimate using a random sample from New York City. Assume that there is more residential stratification by income in New York than in San Francisco. How would you expect the intercept and slope coefficients to differ across the two regression fits?