Midterm

Ec240a – Second Half, Fall 2020

The exam is open book, however you are not allowed to consult with classmates or other people (including by, for example, posting questions on Stack Exchange or other online discussion boards). You may consult your course notes, problem sets, books, articles, material on web etc.

[1] **[20 points]** You observe a simple random sample of size N from the population

$$Y_0 \sim N\left(\mu, \sigma^2\right)$$

as well as a second, independent, simple random sample, also of size N, from the population

$$Y_1 \sim N\left(\mu, 4\sigma^2\right).$$

The value of σ^2 is known. Consider the family of estimates of μ

$$\hat{\mu}\left(c_0, c_1\right) = c_0 \bar{Y}_0 + c_1 \bar{Y}_1,$$

where $\bar{Y}_0 = \frac{1}{N}\sum_{i=1}^{N} Y_{0i}$ and $\bar{Y}_1 = \frac{1}{N}\sum_{i=1}^{N} Y_{1i}$.

[a] **[5 points]** Show that mean squared error equals

$$\mathbb{E}\left[\left(\hat{\mu}\left(c_0, c_1\right) - \mu\right)^2\right] = \frac{c_0^2\sigma^2}{N} + \frac{c_1^2 4\sigma^2}{N} + \left(1 - c_0 - c_1\right)^2 \mu^2. \tag{1}$$

[b] **[5 points]** Derive the oracle estimator (within the family) which minimizes (1).

[c] **[5 points]** Show that

$$\hat{R}\left(c_0, c_1\right) = \frac{c_0^2\sigma^2}{N} + \frac{c_1^2 4\sigma^2}{N} + \left(1 - c_0 - c_1\right)^2 \frac{1}{2}\left\{\bar{Y}_0^2 + \bar{Y}_1^2 - \frac{5\sigma^2}{N}\right\} \tag{2}$$

is an unbiased estimate of of (1). Can you propose another unbiased risk estimate? Why would you prefer one unbiased risk estimate over another? **[6 sentences].**

[d] **[5 points]** Describe in *words* how one might use (2) to construct an implementable estimator of μ. **[6 sentences].**

[2] **[15 points]** Let Y denote log-earnings and X years of completed schooling for a cohort of workers. Assume a random sample of size N is available from this population. Let $D_x = 1$ if $X = x$ and zero otherwise. Assume that $X \in \{0, \ldots, 16\}$ with positive probability attached to each support point.

[a] **[5 points]** Let

$$\mathbb{E}^*\left[Y \mid D_1, \ldots, D_L\right] = \alpha_0 + \sum_{l=1}^{16} \gamma_{0l} D_l.$$

What is the relationship between this linear predictor and $\mathbb{E}\left[Y \mid X = x\right]$? **[6 sentences].**

[b] **[5 points]** Assume that $\Pr\left(X = 6\right) = 0$. Is the linear predictor defined in part [a] still well-defined? Why or why not? **[6 sentences].**

[c] **[5 points]** You hypothesize that $\mathbb{E}\left[Y \mid X = x\right]$ is linear in x. Consider the linear predictor in part [a] and let $\beta = \left(\alpha, \gamma_1, \ldots, \gamma_{16}\right)'$. Show how your hypothesis may be equivalently expressed as set of linear

restrictions of the form $C\beta_0 = c$. Provide explicit expressions for C and c. Describe in detail how you would construct a test statistic for your hypothesis. What is the asymptotic sampling distribution of your statistic under the null? Assume that your have a consistent estimate $\hat{\Lambda}$ of the asymptotic variance-covariance matrix of $\sqrt{N}\left(\hat{\beta}-\beta\right)$, with $\hat{\beta}$ the least squares estimate. **[10 sentences].**

[3] **[25 Points]** Let Y equal tons of banana's harvested in a given season for a randomly sampled Honduran banana planation. Output is produced using labor and land according to $Y = AL^{\alpha_0}D^{1-\alpha_0}$, where L is the number of employed workers and D is the size of the plantation in acres and we assume that $0 < \alpha_0 < 1$. The price of a unit of output is P, while that of a unit of labor is W. These prices may vary across plantations (e.g., due to transportation costs, labor market segmentation etc.). We will treat D as a fixed factor; A captures sources of plantation-level differences in farm productivity due to unobserved differences in, for example, soil quality and managerial capacity. Plantation owners choose the level of employed labor to maximize profits. The observed values of L are therefore solutions to the optimization problem:

$$L = \arg\max_{l} P \cdot Al^{\alpha_0}D^{1-\alpha_0} - W \cdot l.$$

[a] **[2 Points]** Show that the amount of employed labor is given by

$$L = \left\{\alpha_0 \frac{P}{W} A\right\}^{\frac{1}{1-\alpha_0}} D. \tag{3}$$

[b] **[3 Points]** Let $a_0 = \frac{1}{1-\alpha_0}\ln\alpha_0 + \frac{1}{1-\alpha_0}\mathbb{E}[\ln A]$, $b_0 = \frac{1}{1-\alpha_0}$, and $V = \frac{1}{1-\alpha_0}\{\ln A - \mathbb{E}[\ln A]\}$. Show that the log of the labor-land ratio is given by

$$\ln\left(\frac{L}{D}\right) = a_0 + b_0 \ln\left(\frac{P}{W}\right) + V \tag{4}$$

and that, letting $c_0 = \mathbb{E}[\ln A]$ and $U = \ln A - \mathbb{E}[\ln A]$, the log of planation yield (output per unit of land) is given by

$$\ln\left(\frac{Y}{D}\right) = c_0 + \alpha_0 \ln\left(\frac{L}{D}\right) + U. \tag{5}$$

[c] **[5 Points]** Briefly discuss the content and plausibility of the restriction. **[4 to 6 sentences].**

$$\mathbb{E}[\ln A|\ln(P/W)] = \mathbb{E}[\ln A]. \tag{6}$$

[d] **[5 Points]** Using (4), (5) and (6) show that the coefficient on $\ln(L/D)$ in $\mathbb{E}^*[\ln(Y/D)|\ln(L/D)]$ equals

$$\alpha_0 + (1-\alpha_0)\frac{\mathbb{V}(\ln A)}{\mathbb{V}(\ln A) + \mathbb{V}(\ln(P/W))}.$$

Provide some economic intuition for this result. **[4 to 6 sentences]** .

[e] **[5 Points]** Using (4), (5) and (6) show that the coefficient on $\ln(L/D)$ in $\mathbb{E}^*[\ln(Y/D)|\ln(L/D), V]$ equals α_0. Provide some economic intuition for this result. **[4 to 6 sentences]**

[f] **[5 Points]** Assume that all plantations face the same output price (P) and labor cost (W). What value does the coefficient on $\ln(L/D)$ in $\mathbb{E}^*[\ln(Y/D)|\ln(L/D)]$ equal now? Why? **[4 to 6 sentences].**

[4] **[25 Points]** Let W, X be a pair of regressors with the property that $\mathbb{C}(W, X) = 0$. Show that, for

outcome, Y,

$$\mathbb{E}^*\left[Y|\, W, X\right] = \mathbb{E}^*\left[Y|\, W\right] + \mathbb{E}^*\left[Y|\, X\right] - \mathbb{E}\left[Y\right]. \tag{7}$$

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

[a] **[3 Points]** First show that

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

[b] **[5 Points]** Second verify (7) result using the Projection Theorem.

[c] **[2 Points]** Next show that

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

[d] **[5 Points]** The Vice Chancellor for Undergraduate Education is interested in boosting academic performance among first year students. She random 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] **[5 Points]** 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. **[6 to 10 sentences].**

[f] **[5 Points]** Outline a more informative experiment for the Vice Chancellor. Explain why is it is “better” than the experiment described in part [d]. **[6 to 10 sentences].**

<u>[5]</u> **[15 Points]** Let $X \sim \text{Uniform}\left[-1, 1\right]$ and assume that

$$Y = -\frac{2}{3} + X^2 + V, \;\; V|\, X \sim \mathcal{N}\left(0, \sigma^2\right) \tag{8}$$

[a] **[2 Points]** Calculate $\mathbb{E}\left[Y|\, X\right]$

[b] **[2 Points]** Calculate $\mathbb{E}\left[X^2\right]$ and $\mathbb{V}\left(X\right)$

[c] **[2 Points]** Calculate $\mathbb{E}\left[Y\right]$

[d] **[2 Points]** Calculate $\mathbb{C}\left(X, Y\right)$ and also the coefficient on X in $\mathbb{E}^*\left[Y|\, X\right]$.

[e] **[2 Points]** Let $U = Y - \mathbb{E}^*\left[Y|\, X\right]$. Show that $\mathbb{C}\left(U, X\right) = 0$. Give an intuitive explanation for this result.

[f] **[2.5 Points]** Find a function $g\left(X\right)$ such that that $\mathbb{C}\left(U, g\left(X\right)\right) = \mathbb{V}\left(X^2\right)$. Give an intuitive explanation for your answer. **[2 to 4 sentences].**

[g] **[2.5 Points]** Describe how your answers in (d) to (f) would change if (8) held but now $X \sim \text{Uniform}\left[0, 2\right]$. You may find it helpful to sketch a figure. **[2 to 4 sentences].**