First Midterm

Ec142 – Spring 2017

Please read each question carefully. Start each question on a new bluebook page. The use of calculators and other computational aides is not allowed. Good luck!

[1] **[5 Points]** Please write your full name on this exam sheet and turn it in with your bluebook.

[2] **[30 Points]** The Vice Chancellor for Undergraduate Education is concerned about students dropping out for Cal prior to finishing the requirements for a BA. She provides you with the following Table. The table refers to the Cal students who first arrived on campus in the Fall semester of 2013.

	Number in F13 Still at Cal	Number Dropping out	Number Transferring	Hazard	Survival	Std. Error
F13	6,000	1500	200			
S14		1530	70			
F14		940	260			
S15		350	150			

The "Number Transferring" column reports the number of students who transfer to another University at the close of the semester. You may assume that these students are lost to further follow-up. The "Std. Error" column refers to the standard error of the survival function.

[a] **[5 Points]** State and discuss the "random censoring" assumption introduced in lecture. Is this assumption credible in the current context? Explain.

[b] **[10 Points]** Under the maintained assumption of random censoring fill-in the empty cells in the table. What is the median number of semesters enrolled at Cal prior to drop-out.

[c] **[10 Points]** The Vice Chancellor provides you with additional information on whether a student is a "first generation" college student. She is concerned that dropout behavior may vary across first generation and non first-generation students. Explain, in detail, how you would conduct a discrete hazard analysis targeted toward this question for the Vice Chancellor.

[d] **[5 Points]** Discuss how you could implement a model for part (c) with "constant baseline hazard". Is such a model reasonable in the current setting?

[2] **[30 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

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

[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

$$\ln\left(\frac{L}{D}\right) = a_0 + b_0 \ln\left(\frac{P}{W}\right) + V \tag{2}$$

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

$$\ln\left(\frac{Y}{D}\right) = c_0 + \alpha_0 \ln\left(\frac{L}{D}\right) + U. \tag{3}$$

[c] **[5 Points]** Briefly discuss the content and plausibility of the restriction

$$\mathbb{E}[\ln A | \ln(P/W)] = \mathbb{E}[\ln A]. \tag{4}$$

[d] **[10 Points]** Using (2), (3) and (4) 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.

[e] **[5 Points]** Using (2), (3) and (4) 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.

[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?

[3] **[35 Points]** For a random draw from the population of US workers, let Y equal log earnings and X be a binary indicator taking a value of one if the worker is female and zero otherwise. Let $\{(Y_i, X_i)\}_{i=1}^N$ be a random sample of size N. Let N_1 denote the number of sampled units that are women (i.e., $X = 1$) and $N_0 = N - N_1$ the number that are male. Assume that

$$Y_i = \alpha_0 + \beta_0 X_i + U_i$$

with

$$Q_{U|X}(1/2|X) = 0.$$

Let a and b be a candidate values for 'the truth' (i.e., α_0 and β_0). Let $u^1_{1/2}(a,b)$ be the median of $U(a,b) = Y - a - bX$ given $X = 1$. Let $u^0_{1/2}(a,b)$ be the corresponding median given $X = 0$. Let $R_1(a,b), \ldots, R_{N_1}(a,b)$ denote the N_1 order statistics of $U(a,b)$ in the $X_i = 1$ subsample. Let $S_1(a,b), \ldots, S_{N_0}(a,b)$ denote the N_0 corresponding statistics from the $X_i = 0$ subsample.

[a] **[5 Points]** Interpret (in words) the parameters α_0 and β_0. What is true about the distribution of male versus female earnings if $\beta_0 = 0$?

[b] **[2 Points]** What is the median of $U(\alpha_0, \beta_0)$ given, respectively, $X = 1$ and $X = 0$?

[c] **[3 Points]** Assume that $a = \alpha_0$ and $b = \beta_0$. Let $j/(N_1+1) < 1/2 \leq (j+1)/(N_1+1)$. Before looking at your sample you are asked to guess the value of $(R_j(a,b) + R_{j+1}(a,b))/2$. What is your guess? Justify your answer.

[d] **[5 Points]** Let $N_1 = 4$ and $N_0 = 4$ (for this part of the problem only). Consider the order statistic intervals $[R_1(a,b), R_4(a,b)]$ and $[S_1(a,b), S_4(a,b)]$. Assume $a = \alpha_0$ and $b = \beta_0$; what is the ex ante probability that each of these intervals contain zero? Be sure to explain your work.

[e] **[5 Points]** Let a and b be some candidate intercept and slope values. Describe, in detail, an estimate of $u^0_{1/2}(a,b)$ and $u^1_{1/2}(a,b)$? Denote these estimates by, respectively, $\hat{u}^1_{1/2}(a,b)$ and $\hat{u}^0_{1/2}(a,b)$.

[f] **[5 Points]** Describe how to construct an approximate 95 percent confidence interval for $u^0_{1/2}(a,b)$ and $u^1_{1/2}(a,b)$?

[g] **[5 Points]** Describe how to construct an estimate of the asymptotic sampling variances of $\sqrt{N}\left(\hat{u}^1_{1/2}(a,b) - u^1_{1/2}(a,b)\right)$ and $\sqrt{N}\left(\hat{u}^0_{1/2}(a,b) - u^0_{1/2}(a,b)\right)$?

[h] **[5 Points]** Using your estimates from part (e) and sampling variance from part (g)

sketch a procedure for testing the joint null hypothesis $H_0 : \alpha_0 = a, \beta_0 = b$.