Ec142, Spring 2018

Professor Bryan Graham

Review Sheet 1

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 midterm exam will occur in class on Thursday, March 22nd.

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

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

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

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

[b] Second verify the result using the Projection Theorem.

[c] Finally show that

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

[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].

[2] The Vice Chancellor for Undergraduate Education is concerned about students dropping out of 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 still at Cal	Number Dropping out	Number Transferring	Hazard	Survival	Std. Error
F13	6,000	500	200			
S14		530	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] State and discuss the "random censoring" assumption introduced in lecture. Is this assumption credible in the current context? Explain.

[b] 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] 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.

[3] 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 $(\delta_0^* W + V^*) s + \frac{\kappa}{2} s^2$. Here W is an observable variable which shifts the marginal cost of schooling and V^* is unobserved heterogeneity. You may assume that both U and V^* are conditionally mean zero given W. Agents choose years of completed schooling to maximize expected utility

$$S = \arg\max_{s \in \mathbb{S}} \mathbb{E}\left[\log Y(s) - (\delta_0^* W + V^*) s - \frac{\kappa}{2} s^2 \middle| W, V \right].$$

[a] Show that observed schooling is given by

$$S = \gamma_0 + \delta_0 W + V, \quad \mathbb{E}[V | W] = 0$$

for $\gamma_0 = \beta_0/\kappa$, $\delta_0 = -\delta^*/\kappa$, and $V = -V^*/\kappa$.

[b] Assume that W measures commute time to the closest four year college from a respondent's home during adolescence. What sign do you expect δ_0 to have? Explain.

[c] Assume that $\mathbb{E}[U | W, V] = \mathbb{E}[U | V] = \lambda V$. Restate this assumption in words (HINT: Think about V as a latent variable/attribute). What sign do you expect λ to have? Briefly argue for and against this assumption?

[d] Let $\log Y = \log Y(S)$ denote actual earnings. Show that

$$\mathbb{E}^* [\log Y | S, V] = \alpha_0 + \beta_0 S + \lambda V. \tag{1}$$

[e] What determines variation in S conditional on $V = v$? What is the relationship between this variation and the unobserved determinants of log earnings? Use your answers to provide an intuitive explanation (i.e., use words) for why the coefficient on schooling in (1) equals β_0.

[f] The random sample $\{(Y_i, S_i, W_i)\}_{i=1}^N$ is available. Suggest a procedure for consistently estimating β_0.

[g] Let

$$\mathbb{E}^* [\log Y | S] = a_0 + b_0 S.$$

From you analysis in part [f] you learn that $\lambda \approx 0$. Guess what value b_0 takes. Justify your answer.

[4] You've been hired by the Government of Honduras to assess the efficacy of treatment for decompression sickness among lobster divers in La Moskitia. In this region of Honduras lobsters are harvested by divers who, on occasion, get decompression sickness which may result in partial paralysis or worse. You are provided the following table of information about 300 diving accident victims.

		$Y = 0$ (No Limp)	$Y = 1$ (Limp)
$X = 0$ (Untreated)	$W = 0$ (Depth $<$ 75')	90	10
	$W = 1$ (Depth $\geq$ 75')	10	40
$X = 1$ (Treated)	$W = 0$ (Depth $<$ 75')	30	20
	$W = 1$ (Depth $\geq$ 75')	50	50

[a] What is the probability of a victim walking with a limp conditional on treatment ($X = 1$) and non-treatment ($X = 0$)?

[b] What is the probability of a victim receiving treatment conditional on having dived "deep" ($W = 1$) vs. "shallow" ($W = 0$)?

[c] A government official worries that treatment is harming the divers and thinks it would be better to do nothing. Present a counter-argument to this official.

[d] Let $Y(0)$ and $Y(1)$ denote a divers potential outcome given non-treatment and treatment respectively. Discuss the conditional independence assumption assumption

$$(Y(0), Y(1)) \perp X | W = 0, 1.$$

Make a positive and negative argument for this assumption.

[e] Using the assumption in part [d] construct the IPW estimate of the average treatment effect on the outcome. Report your result to the government official. Your report should include an explanation for why and how your are adjusting for accident depth. Is treatment effective?

[f] Say instead you were given the table:

		$Y = 0$ (No Limp)	$Y = 1$ (Limp)
$X = 0$ (Untreated)	$W = 0$ (Depth $<$ 75')	90	10
	$W = 1$ (Depth $\geq$ 75')	0	0
$X = 1$ (Treated)	$W = 0$ (Depth $<$ 75')	30	20
	$W = 1$ (Depth $\geq$ 75')	75	75

Can you compute the ATE is this case? Why or why not?

[5] 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

$$Q_{U|X}(\tau | X) = 0.$$

Let a and b be a candidate values for 'the truth' (i.e., α_0 and β_0). Let $u_\tau^1(a,b)$ be the τ quantile of of $U(a,b) = Y - a - bX$ given $X = 1$. Let $u_\tau^0(a,b)$ be the corresponding τ quantile 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] 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] What is the τ quantile of of $U(\alpha_0, \beta_0)$ given, respectively, $X = 1$ and $X = 0$?

[c] Assume that $a = \alpha_0$ and $b = \beta_0$. Let $j/(N_1+1) < \tau \leq (j+1)/(N_1+1)$. 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] Assume that $a = \alpha_0$ and $b = \beta_0$. Let $j/(N_1+1) < \tau - \epsilon \leq (j+1)/(N_1+1)$ for $0 < \epsilon < \tau$. Before looking at your sample you are asked to guess the *sign* of $(R_j(a,b) + R_{j+1}(a,b))/2$. What is your guess? Justify your answer.

[e] Let $\tau = 1/2$, $N_1 = 3$, and $N_0 = 3$ (for this part of the problem only). Consider the order statistic intervals $[R_1(a,b), R_3(a,b)]$ and $[S_1(a,b), S_3(a,b)]$. 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.

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

[g] Describe how to construct an approximate 95 percent confidence interval for $u_{1/2}^0(a,b)$ and $u_{1/2}^1(a,b)$?

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

[i] Using your estimates from part (f) and sampling variance from part (h) sketch a procedure for testing the joint null hypothesis $H_0 : \alpha_0 = a, \beta_0 = b$.

[6] 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] Show that the amount of employed labor is given by

$$L = \left\{\alpha_0 \frac{P}{W} A\right\}^{\frac{1}{1-\alpha_0}} D. \tag{2}$$

[b] 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

$$\ln\left(\frac{L}{D}\right) = a_0 + b_0 \ln\left(\frac{P}{W}\right) + V \tag{3}$$

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

$$\ln\left(\frac{Y}{D}\right) = c_0 + \alpha_0 \ln\left(\frac{L}{D}\right) + U. \tag{4}$$

[c] Briefly discuss the content and plausibility of the restriction

$$\mathbb{E}\left[\ln A \right| \ln(P/W)] = \mathbb{E}[\ln A]. \tag{5}$$

[d] Using (3), (4) and (5) show that the coefficient on $\ln(L/D)$ in $\mathbb{E}^*\left[\ln(Y/D)\right|\ln(L/D)]$ 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] Using (3), (4) and (5) show that the coefficient on $\ln(L/D)$ in $\mathbb{E}^*\left[\ln(Y/D)\right|\ln(L/D), V]$ equals α_0. Provide some economic intuition for this result.

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