Ec141, Spring 2020

Professor Bryan Graham

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 (indeed I encourage you to do so and also to be generous with one another as you prepare – make use of Zoom, Facebook and other technologies for virtual interaction). This midterm is open book, but – unlike the review sheet – you must work on the midterm without consulting with classmates or any other persons. I have no way of enforcing this rule, but I hope you will all take it seriously in the interest of fairness to everyone. The midterm exam will be posted at 11AM on Thursday 4/30 on bCourses. It will be due by 11AM on Saturday 5/2, also on bCourses. This deadline is firm. Please plan accordingly. Turn in your exam as a pdf document, whether your answers are typewritten or a clean scan of neatly hand-written ones.

[1] 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 (ATE) 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?

[2] You are given a random sample from South Africa in the late 1980s. Each record in this sample includes, Y, an individual's log income at age 40, X the log permanent income of their parents, and D a binary indicator equaling 1 if the respondent is White and zero if they are Black. Let the best linear predictor of own log income at age forty given parents' log permanent income and own race be

$$\mathbb{E}^*\left[Y \middle| X, D\right] = \alpha_0 + \beta_0 X + \gamma_0 D.$$

[a] Let $Q = \Pr\left(D = 1\right)$, assume that $\mathbb{V}\left(X \middle| D = 1\right) = \mathbb{V}\left(X \middle| D = 0\right) = \sigma^2$ and recall the analysis of variance formula $\mathbb{V}\left(X\right) = \mathbb{V}\left(\mathbb{E}\left[X \middle| D\right]\right) + \mathbb{E}\left[\mathbb{V}\left(X \middle| D\right)\right]$. Show that

$$\mathbb{V}\left(X\right) = Q\left(1 - Q\right)\left\{\mathbb{E}\left[X \middle| D = 1\right] - \mathbb{E}\left[X \middle| D = 0\right]\right\}^2 + \sigma^2.$$

[b] Let $\mathbb{E}^*\left[D \middle| X\right] = \kappa + \lambda X$. Show that

$$\lambda = \frac{Q\left(1 - Q\right)\left\{\mathbb{E}\left[X \middle| D = 1\right] - \mathbb{E}\left[X \middle| D = 0\right]\right\}}{Q\left(1 - Q\right)\left\{\mathbb{E}\left[X \middle| D = 1\right] - \mathbb{E}\left[X \middle| D = 0\right]\right\}^2 + \sigma^2}.$$

[c] Assume that $\beta_0 = 0$. Show that in this case $\gamma_0 = \mathbb{E}\left[Y \middle| D = 1\right] - \mathbb{E}\left[Y \middle| D = 0\right]$.

[d] Let $\mathbb{E}^*\left[Y \middle| X\right] = a + bX$. Maintaining the assumption that $\beta_0 = 0$ show that

$$b = \frac{Q\left(1 - Q\right)\left\{\mathbb{E}\left[Y \middle| D = 1\right] - \mathbb{E}\left[Y \middle| D = 0\right]\right\}\left\{\mathbb{E}\left[X \middle| D = 1\right] - \mathbb{E}\left[X \middle| D = 0\right]\right\}}{Q\left(1 - Q\right)\left\{\mathbb{E}\left[X \middle| D = 1\right] - \mathbb{E}\left[X \middle| D = 0\right]\right\}^2 + \sigma^2}.$$

[e] Let $Q\left(1 - Q\right) = 1/10$, $\sigma^2 = 3/10$ and $\mathbb{E}\left[Y \middle| D = 1\right] - \mathbb{E}\left[Y \middle| D = 0\right] = \mathbb{E}\left[X \middle| D = 1\right] - \mathbb{E}\left[X \middle| D = 0\right] = 3$. Provide a numerical value for $\mathbb{V}\left(X\right)$ and b.

[f] On the basis of β_0 a member of the National Party argues that South Africa is a highly mobile society. One the basis of b a member of the African National Congress argues that it is a highly immobile one. Comment on the relative merits of these two assertions.

[3] Let $C_t = 1$ if an individual (child) went to college and zero otherwise. Let $C_{t-1} = 1$ if the corresponding individual's parent went to college and zero otherwise. The following table gives the joint distribution of father and sons' college attendance:

	$C_t = 0$	$C_t = 1$
$C_{t-1} = 0$	0.60	0.20
$C_{t-1} = 1$	0.10	0.10

For example 20% percent of the population consists of pairs with a father who did not attend college, but a son who did.

[a] Among children of college graduates, what fraction go on to complete college themselves? Among children of non-graduates, what fraction go on to complete college themselves?

[b] Let $\mathbb{E}^*\left[C_t|\, C_{t-1}\right] = a + bC_{t-1}$; calculate a and b.

[c] The following table gives child's adult earnings, Y_t, for each of the four subpopulations introduced above

	$C_t = 0$	$C_t = 1$	
$C_{t-1} = 0$	\$8,000	\$60,000	.
$C_{t-1} = 1$	\$14,000	\$30,000	

What is the average earnings level of college graduates in this economy? What is the average earnings of non-college graduates? What is the overall average earnings level? Express your answers symbolically using the notation of (conditional) expectations and also provide a numerical answer.

[d] Let $\pi_{c_{t-1}} = \Pr\left(C_{t-1} = c_{t-1}|\, C_t = 1\right)$. Consider the estimand

$$\beta = \sum_{c_{t-1}=0,1} \left\{\mathbb{E}\left[Y|\, C_t = 1, C_{t-1} = c_{t-1}\right] - \mathbb{E}\left[Y|\, C_t = 0, C_{t-1} = c_{t-1}\right]\right\} \pi_{c_{t-1}}.$$

In what sense does β adjust for "covariate differences" between college and non-college graduates? Evaluate β and compare your numerical answer with the raw college - non-college earnings gap you calculated in part (c). Why are these two numbers different?

[e] Gavin Newsom is considering a community college expansion policy. You have been tasked to asked to predict the earnings gain associated with completing a college degree. Gavin estimates that after the community college expansion the distribution of college attendance in California will look like

	$C_t = 0$	$C_t = 1$	
$C_{t-1} = 0$	0.40	0.40	.
$C_{t-1} = 1$	0.05	0.15	

Calculate average earnings in this new economy (you may assume that the mapping from background and education into earnings introduced in part (c) remains the same)? Assume a state tax rate of 10 percent. What is the long run predicted increase in annual tax revenue from the community college expansion? Treat this revenue as a perpetuity and assume a discount rate of 0.05. What is the present value of the increase in tax revenue that is expected to be generated by the community college expansion?

[4] 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 $\left(\delta_0^* W + V^*\right) s + \frac{\kappa}{2} s^2$. 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

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

[a] Show that observed schooling is given by

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

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

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

[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.

[5] 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:

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

[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

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

[d] Using (3), (4) and (5) show that the coefficient on $\ln\left(L/D\right)$ in $\mathbb{E}^*\left[\ln\left(Y/D\right)\right|\ln\left(L/D\right)]$ equals

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

Provide some economic intuition for this result.

[e] Using (3), (4) and (5) show that the coefficient on $\ln\left(L/D\right)$ in $\mathbb{E}^*\left[\ln\left(Y/D\right)\right|\ln\left(L/D\right),V]$ 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 $\ln\left(L/D\right)$ in $\mathbb{E}^*\left[\ln\left(Y/D\right)\right|\ln\left(L/D\right)]$ equal now? Why?

[6] Consider a population of high school graduates. Let Y_1 denote the earnings an individual in this population would get if they completed at least four years of college, let Y_0 denote the earnings they would get if they did not complete college. Let $D = 1$ in an individual actually completes college and zero otherwise. Let Y denote observed earnings which, given the data structure outlined above, equals

$$Y = \left(1-D\right)Y_0 + DY_1.$$

Assume that

$$\left(Y_1, Y_0\right) \perp D|\, X$$

where X is a characteristic measured at the completion of high school but prior to any college attendance. Further assume that

$$\begin{aligned}\mathbb{E}\left[Y_1\right|X] &= \alpha_1 + \gamma_1 X\\ \mathbb{E}\left[Y_0\right|X] &= \alpha_0 + \gamma_0 X.\end{aligned}$$

[a] Across subpopulations homogenous in $X = x$ can we use Y_1 or Y_0 to predict college attendance? Would these variables be informative about college attendance unconditional on X? Can we use an individual's observed wage Y to predict whether they went to college? [4 to 6 sentences]

[b] Let $\beta_0 = E\left[Y_1 - Y_0\right|D = 1]$. Interpret this object. Derive a representation of it in terms of α_0, α_1, γ_0, γ_1 and the distribution of X.

[c] Show that

$$\mathbb{E}\left[Y\right|X, D] = \alpha_0 + \gamma_0 X + \left(\alpha_1-\alpha_0\right)D + \left(\gamma_1-\gamma_0\right)DX.$$

[d] Consider the following least squares fit of log earnings on a dummy for completion of an undergraduate degree using a sample of 1,754 white males from the NLSY79.

$$\text{LogEarnings} = \underset{(0.0220)}{10.0332} + \underset{(0.0357)}{0.4879}\ \text{UNDERGRAD}\ . \tag{6}$$

Now consider the least squares fit which additionally includes an individual's AFQT percentile score and its interaction with UNDERGRAD:

$$\begin{aligned}\text{LogEarnings} = \underset{(0.0573)}{9.8231} &+ \underset{(0.0010)}{0.0040}\ \text{AFQT} \\ &+ \underset{(0.1584)}{0.1898}\ \text{UNDERGRAD} + \underset{(0.0020)}{0.0023}\ \text{UNDERGRAD} \times \text{AFQT}\end{aligned} \tag{7}$$

Finally consider the least squares fit of AFQT on a constant and UNDERGRAD:

$$\text{AFQT} = \underset{(0.73)}{52.27} + \underset{(1.05)}{28.35}\ \text{UNDERGRAD}\ . \tag{8}$$

Assume that $X = AFQT$. Using these results compute an estimate of β_0. Justify and explain your calculations. How does your estimate compare with the coefficient on UNDERGRAD in (6)? Comment on an differences and provide an explanation for them (if needed).

[7] 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	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.