Ec143, Spring 2023

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 first midterm exam will occur in class on Thursday, March 16th.

[1] Let $(X_i', C_i, Y_i)'$ denote the i^{th} random draw of a covariate vector, a censoring time and a duration from a target population of interest. Both the censoring time and the duration are measured in discrete time with known support. The econometrician observes

$$Z_i = \min\{Y_i, C_i\}$$

as well as the non-censored indicator

$$D_i = \begin{cases} 1 & \text{if } Y_i \leq C_i \\ 0 & \text{if } Y_i > C_i \end{cases}.$$

Available to the econometrician is the random sample $(X_1', Z_1, D_1)', (X_2', Z_2, D_2)', \ldots, (X_N,' Z_N, D_N)'$.

[a] You are working as a staff econometrician at *Bespoke Data Analytics, Ltd.*, a boutique economic consulting firm located in the Welsh countryside and serving clients worldwide. You've been asked to study time to battery failure in connection with a class action lawsuit against Wicked Good Car Battery Corp. Your boss, Eustace Griffiths, provides you with a graph of

$$S^{\text{obs}}(y) = \Pr(Z > y)$$

for $y \in \mathbb{Y} = \{y_1, \ldots, y_J\}$. Here Z is the observed time to car battery failure (or censoring) in months (the true, uncensored, battery life length is Y). Show that

$$S^{\text{obs}}(y) = \Pr(Y \geq y, C \geq y).$$

[b] Show that if $Y_i \perp C_i$, then

$$S(y) = \frac{S^{\text{obs}}(y)}{\Pr(C \geq y)}$$

where $S(y) = \Pr(Y > y)$ is the population survival function. How does naively ignoring censoring shape one's conclusions about the distribution of battery life?

[c] One of the research associates at Bespoke produces the following life table (written in beautiful fountain pen). Complete columns 1, 5 and 6 of the table.

(1)	(2)	(3)	(4)	(5)	(6)
Month	'At Risk'	'Number of failures'	'Lost to follow-up'	Hazard Rate	Survival Function
$Y_{(k)}$ (or y)	$N_{(k)}$	$N_{(k)}^{\mathrm{d}}$	$N_{(k)}^{\mathrm{c}}$	$\lambda(y)$	$S(y)$
1	1000	250	150		
2		150	0		
3		50	100		
4		150	100		
5		25	25		

[d] Using the life table from part [c] above construct an estimate of the median car battery life. What fraction of batteries do you estimate will last more than five months?

[e] Wicked Good Car Battery Corp. has two manufacturing facilities. Let $X \in \{0, 1\}$ indicate whether a battery was made in the Pitcairn Island or South Georgia Island facility. Another research associate at Bespoke produces the follow data table (written on high stock graph paper in fountain pen)

i	X	Z	D
1	1	4	1
2	1	3	0
3	0	1	0
4	0	2	1
$\vdots$			

Let $W_{iy} = 1$ if $Z_i = y$ and $D_i = 1$ and zero otherwise. Construct the first four observation's contributions to the "person period" dataset and place them in the table below.

i	y	W_{iy}	X_i

[f] Assume that the conditional hazard function for battery failure equals

$$\lambda\left(y|\, X;\theta\right) = \Pr\left(Y = y|\, Y \geq y, X\right) = \frac{\exp\left(\gamma_0 + \gamma_1 y + X\beta\right)}{1 + \exp\left(\gamma_0 + \gamma_1 y + X\beta\right)}. \tag{1}$$

for some some $\theta = \theta_0 \in \Theta \subset \mathbb{R}^3$ (here $\theta = (\gamma_0, \gamma_1, \beta)'$. Describe how to you could compute the maximum likelihood estimate of θ_0 using the person-period dataset and a logistic regression program. How does the above specification restrict the "baseline" hazard function? How might you relax this restriction?

[g] Let

$$\frac{\frac{\Pr(Y=y|Y\geq y,X=1)}{1-\Pr(Y=y|Y\geq y,X=1)}}{\frac{\Pr(Y=y|Y\geq y,X=0)}{1-\Pr(Y=y|Y\geq y,X=0)}}$$

be the odds ratio (OR) of failure among units made in the South Georgia Island versus those made in Pitcairn Island. Interpret this expression. Assume that $\hat{\beta} = \ln 2$. What is the estimated OR? Jimmy Cook Jr., manager of the South Georgia Island manufacturing facility claims his batteries are among the best in the world. Assess Jimmy's claim.

[2] Prior to 1994 colleges and universities in the United States were exempt from laws prohibiting mandatory retirement, consequently many institutions forced faculty to retire at age 70. After 1994 mandatory retirement rules were prohibited by Congress. Ashenfelter and Card (*AER*, 2002) study the effects of this exemption expiration on faculty retirement behavior in a sample of 104 colleges and universities. Let Y equal age of retirement, C equal the age at which a faculty-member is lost to follow-up, $D = \mathbf{1}\left(Y \leq C\right)$ be a censoring indicator, and $Z = \min\left(Y, C\right)$ be the observed age at exit from the sample. Let T_y (for $y = 60, 61\ldots, 72, 73$) equal the calendar year during which an individual was age y. So, for

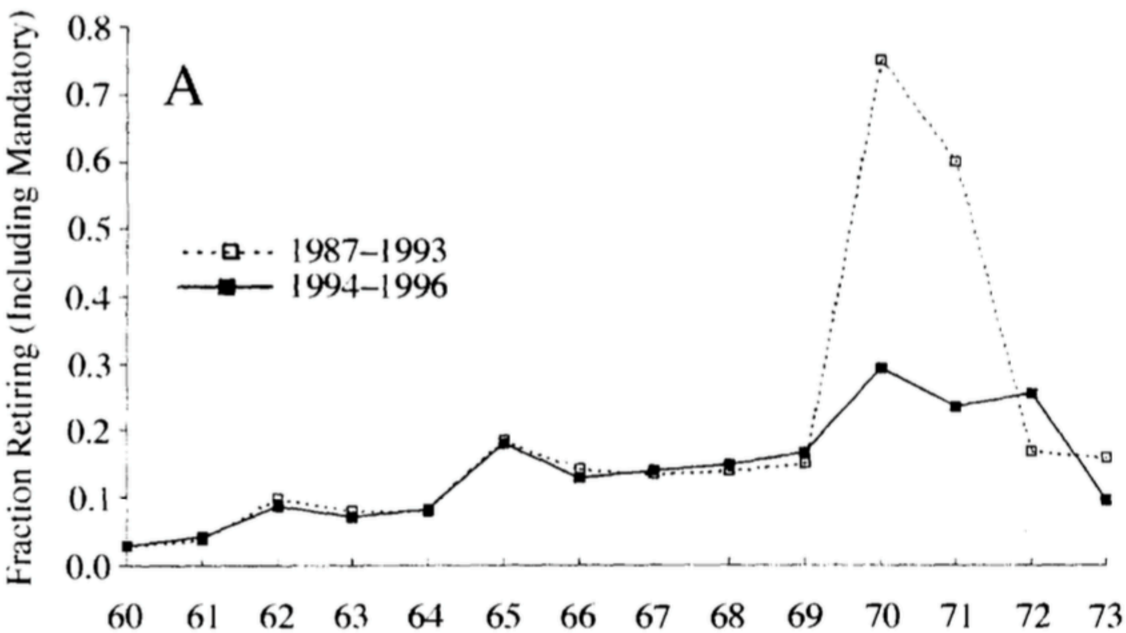
example, an individual who turned sixty in 1992 would have $T_{60} = 1992$, while one who did so in 1999 would have $T_{60} = 1999$.

[a] Let $X = 0$ if $T_{70} < 1994$ and $X = 1$ if $T_{70} \geq 1994$. Assume that

$$\lambda(y|X;\theta) = \Pr(Y = y|Y \geq y, X) = \frac{\exp(\gamma_y + X\beta_0 + X \times \mathbf{1}(y \geq 70)\beta_1)}{1 + \exp(\gamma_y + X\beta_0 + X \times \mathbf{1}(y \geq 70)\beta_1)}. \quad (2)$$

for some some $\theta = \theta_0 \in \Theta \subset \mathbb{R}^{16}$ (here $\theta = (\gamma_{60}, \gamma_{62}, \ldots, \gamma_{73}, \beta_0, \beta_1)'$. In the context of the Ashenfelter and Card (*AER*, 2002) study interpret the hazard function $\lambda(y|X;\theta)$ when $X = 0$ and when $X = 1$. Provide an interpretation of β_0 and β_1.

[b] Reference the figure below when answering the following questions (justify your answers). What signs do you expect β_0 and $\beta_1$to take? Does the evidence appear consistent with the hypothesis that $\beta_0 = 0$ and $\beta_1 < 0$?



[c] Assume that $D \perp Y|X$. Interpret this assumption. Describe how it could be violated.

[d] Assume the first four lines of the Ashenfelter and Card (*AER*, 2002) dataset equal

	Z	D	X
1	65	0	0
2	72	1	0
3	61	1	1
4	70	0	1

What are these units' contributions to the corresponding "person-period" dataset? Write out the corresponding rows. Describe, in detail, how you could use this person period dataset to construct estimates of $\gamma_{60}, \gamma_{61}, \ldots, \gamma_{73}$, β_0 and β_1 .

[e] Let $S\left(y|\, X\right) = \Pr\left(Y > y|\, X\right)$. Does $\Pr\left(Z > y|\, X\right) = S\left(y|\, X\right)$? If not, does $\Pr\left(Z > y|\, X\right) > S\left(y|\, X\right)$ or $\Pr\left(Z > y|\, X\right) < S\left(y|\, X\right)$? Why? Describe a method for constructing an estimate of $S\left(y|\, X\right)$. Describe, in detail, how you could use this estimate to compute the effect of the end of mandatory retirement on median retirement age. Use the information in the table below to implement your procedure. Specifically construct an estimate of $S\left(y|\, X = 1\right)$ and $S\left(y|\, X = 0\right)$ for $y = 60, 61...., 72$. Use your results to construct an estimate of the median retirement age before and after 1994. Comment.

TABLE 2—AGE-SPECIFIC RETIREMENT RATES, BEFORE AND AFTER 1994

			Average retirement rate		Change in retirement rate	
Age	Number of observations	Percentage post-1994	1987–1993	1994–1996	Unadjusted	Adjusted from logit
60	7,343	31.8	3.3	3.0	−0.3	−0.2
			(0.3)	(0.4)	(0.4)	(0.5)
61	7,027	32.4	4.1	4.4	0.3	0.3
			(0.3)	(0.4)	(0.5)	(0.5)
62	6,665	32.9	10.3	8.9	−1.4	−1.4
			(0.5)	(0.6)	(0.8)	(0.8)
63	5,838	34.5	8.5	7.3	−1.3	−1.1
			(0.5)	(0.6)	(0.7)	(0.8)
64	5,222	35.4	8.4	8.5	0.1	0.1
			(0.5)	(0.7)	(0.8)	(0.8)
65	4,650	35.1	19.3	18.1	−1.2	−1.4
			(0.7)	(1.0)	(1.2)	(1.3)
66	3,653	35.1	14.7	13.0	−1.7	−1.9
			(0.7)	(0.9)	(1.2)	(1.3)
67	2,969	34.2	13.8	14.0	0.1	−0.1
			(0.8)	(1.1)	(1.3)	(1.4)
68	2,453	34.2	14.3	14.6	0.4	0.7
			(0.9)	(1.2)	(1.5)	(1.5)
69	2,004	33.7	15.4	16.7	1.3	0.6
			(1.0)	(1.4)	(1.7)	(1.7)
70	1,598	35.1	75.6	29.1	−46.5	−43.7
			(1.3)	(2.0)	(2.4)	(2.5)
71	502	58.6	60.6	23.8	−36.8	−32.2
			(3.4)	(2.5)	(4.2)	(4.0)
72	182	67.0	16.7	25.4	8.7	−3.7
			(4.9)	(4.0)	(6.3)	(7.2)

Notes: Retirement rates expressed as percent per year. Estimated standard errors are in parentheses. An individual's retirement age is measured as of September 1 following the date of retirement. The adjusted change in retirement rates is the normalized regression coefficient from a logit model for the event of retirement, fit by age and including a total of 19 covariates: gender, Ph.D., nonwhite race, region (three dummies), Carnegie classification and public/private status of institution, and six department dummies.