Ec143, Spring 2024

Professor Bryan Graham

Midterm Review Sheet #3

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 second midterm exam will occur in class on Thursday, April 55th.

Part I: Model selection and empirical risk minimization

[1] You have been hired by UNICEF to estimate the prevalence of childhood stunting (low height-for-age) across municipalities in a country where childhood malnutrition is commonplace. Let Y_{it} be the height-for-age Z score of individual $t = 1, \ldots, T$ in municipality $i = 1, \ldots, N$. In each municipality you draw T children at random and compute the average height-for-age Z score

$$\bar{Y}_i = \frac{1}{T} \sum_{t=1}^{T} Y_{it}.$$

You assume that $Y_{it} | \theta_i \sim \mathcal{N}(\theta_i, \sigma^2)$ for $i = 1, \ldots, N$ and $t = 1, \ldots, T$. In this model the expected height-for-age Z score, θ_i, varies across municipalities. Your goal is to estimate the municipality (population) means $\theta_1, \theta_2, \ldots, \theta_N$. Municipalities with low θ_i estimates will be slated to receive new anti-hunger and nutrition programs. Initially you may assume that σ^2 is known (in a healthy population of children $\sigma^2 \approx 1$ since height-for-age Z scores are calibrated to have unit variance in such a setting).

[a] Explain why, if $f(y_{it} | \theta_i)$ is Gaussian, the municipality mean is also Gaussian:

$$\bar{Y}_i \big| \theta_i \sim \mathcal{N}\left(\theta_i, \frac{\sigma^2}{T}\right).$$

[b] Let $\|\mathbf{m}\| = \left[\sum_{i=1}^{N} m_i^2\right]^{1/2}$ denote the Euclidean norm of a vector. Let $\theta = (\theta_1, \ldots, \theta_N)'$. Show that

$$\mathbb{E}\left[\left\|\hat{\theta} - \theta\right\|^2\right] = \sum_{i=1}^{N} \mathbb{E}\left[\left(\hat{\theta}_i - \theta_i\right)^2\right],$$

with $\hat{\theta}$ some estimate – based upon the sample data $\mathbf{Y} = (Y_{11}, \ldots, Y_{1T}, \ldots, Y_{N1}, \ldots, Y_{NT})'$ – of θ. Explain why this measures *expected* estimation accuracy or *risk*? What is being averaged in the expectation?

[c] Further show that

$$\mathbb{E}\left[\left\|\hat{\theta}-\theta\right\|^2\right]=\sum_{i=1}^{N}\mathbb{V}\left(\hat{\theta}_i\right)+\sum_{i=1}^{N}\left(\mathbb{E}\left[\hat{\theta}_i\right]-\theta_i\right)^2.$$

Interpret this expression.

[d] Consider the following family of estimators for θ_i (for $i=1,\ldots,N$):

$$\hat{\theta}_i=(1-\lambda)\,\bar{Y}_i+\lambda\mu,$$

with μ the country-wide mean of Y_{it} (i.e., the expected height-for-age Z score of a randomly sampled child from the full country-wide population). You may assume that μ is known (perhaps from prior research). Assume that $0\leq\lambda\leq 1$. Interpret this estimator? Why might the estimator with $\lambda=0$ be sensible? How might you justify the estimator when $\lambda>0$.

[e] Show, for the family of estimates introduced in part [d], that

$$\mathbb{E}\left[\left\|\hat{\theta}-\theta\right\|^2\right]=(1-\lambda)^2\,\frac{N}{T}\sigma^2+\lambda^2\sum_{i=1}^{N}(\theta_i-\mu)^2.$$

You hear, in the hallways of Evans, that "small λ means small bias" and "big λ means low variance". Explain?

[f] Show that the risk-minimizing choice of λ, say λ^*, is

$$\lambda^*=\frac{N\sigma^2}{N\sigma^2+\sum_{i=1}^{N}T\,(\theta_i-\mu)^2}.$$

Is an estimator based upon λ^* feasible? Why or why not? What is the optimal choice of λ^* as $T\to\infty$? Provide some intuition for your answer. What happens to the optimal choice of λ^* as σ^2 becomes large? Provide some intuition for your answer.

[g] Show that

$$\sum_{i=1}^{N}\mathbb{E}\left[\left(\bar{Y}_i-\hat{\theta}_i\right)^2\right]=\sum_{i=1}^{N}\mathbb{E}\left[\left(\bar{Y}_i-\theta_i\right)^2\right]+\sum_{i=1}^{N}\mathbb{E}\left[\left(\hat{\theta}_i-\theta_i\right)^2\right]-\frac{2\sigma^2}{T}\mathrm{df}\left(\hat{\theta}\right)$$

with the degree-of-freedom of $\hat{\theta}$ (or *model complexity*) equal to

$$\mathrm{df}\left(\hat{\theta}\right)=\sum_{i=1}^{N}\frac{T}{\sigma^2}\mathbb{C}\left(\bar{Y}_i,\hat{\theta}_i\right).$$

We call the term to the left of the first equality above *apparent error*.

[h] Show that

$$\sum_{i=1}^{N} \mathbb{E}\left[\left(\bar{Y}_i - \theta_i\right)^2\right] = \frac{N}{T}\sigma^2$$

and also, for the family of estimates indexed by λ introduced in part [d] above, that

$$\mathrm{df}\left(\hat{\theta}\right) = N\left(1 - \lambda\right).$$

[i] Calculate apparent error and model complexity for $\hat{\theta}$ when $\lambda = 0$. Explain?

[j] Calculate apparent error and model complexity for $\hat{\theta}$ when $\lambda = 1$. Explain?

[k] You are roaming around Evans Hall looking for Professor Graham's office. You can't find his office because of the confusing floor plan. However, after a few hours of wandering around aimlessly, you bump into someone who introduces himself as Chuck Stein. He rearranges the expression you derived in part [g] above to get

$$\sum_{i=1}^{N} \mathbb{E}\left[\left(\hat{\theta}_i - \theta_i\right)^2\right] = -\sum_{i=1}^{N} \mathbb{E}\left[\left(\bar{Y}_i - \theta_i\right)^2\right] + \sum_{i=1}^{N} \mathbb{E}\left[\left(\bar{Y}_i - \hat{\theta}_i\right)^2\right] + \frac{2\sigma^2}{T}\mathrm{df}\left(\hat{\theta}\right).$$

He then notices your results from part [h] further imply that

$$\sum_{i=1}^{N} \mathbb{E}\left[\left(\hat{\theta}_i - \theta_i\right)^2\right] = -\frac{N}{T}\sigma^2 + \sum_{i=1}^{N} \mathbb{E}\left[\left(\bar{Y}_i - \hat{\theta}_i\right)^2\right] + 2N\frac{\sigma^2}{T}\left(1 - \lambda\right).$$

Finally he says therefore an unbiased estimate of risk is:

$$\mathrm{SURE}\left(\bar{Y}, \lambda\right) = -\frac{N}{T}\sigma^2 + \sum_{i=1}^{N} \lambda^2\left(\bar{Y}_i - \mu\right)^2 + 2\frac{N}{T}\sigma^2\left(1 - \lambda\right).$$

Provide an explanation for Chuck Stein's claim. Next show that

$$\mathbb{E}\left[\left(\bar{Y}_i - \mu\right)^2\right] = \frac{\sigma^2}{T} + \left(\theta_i - \mu\right)^2$$

and hence that

$$\mathbb{E}\left[\mathrm{SURE}\left(\bar{Y}, \lambda\right)\right] = \sum_{i=1}^{N} \mathbb{E}\left[\left(\hat{\theta}_i - \theta_i\right)^2\right]$$

as implied by Chuck's unbiasedness claim. HINT: Don't forget the work you've done in part [e] above (and you may need to factor a quadratic equation in λ). Why is this result awesome?

[l] Let $\hat{\lambda}^*$ be the value of λ which minimizes SURE $\left(\bar{Y}, \lambda\right)$. Show that

$$\hat{\lambda}^* = \frac{\frac{N}{T}\sigma^2}{\sum_{i=1}^{N}\left(\bar{Y}_i - \mu\right)^2}.$$

Relate this feasible estimator to the infeasible oracle estimator based upon λ^* defined in part [f] above. HINT: use the expression for $\mathbb{E}\left[\left(\bar{Y}_i - \mu\right)^2\right]$ derived in part [k] to argue that when N is large enough $\hat{\lambda}^* \approx \lambda^*$. Discuss.

[m] You decide to use the estimator based upon $\hat{\lambda}^*$. For each municipality you calculate

$$\hat{\theta}_i = \left(1 - \hat{\lambda}^*\right)\bar{Y}_i + \hat{\lambda}^*\mu,$$

and then report when $\hat{\theta}_i < -1$. You tell the Minister of Health that those municipalities where $\hat{\theta}_i < -1$ should be targeted for supplemental child nutrition programs to combat stunting. A few month's later the mayor of village $i = 19$ comes to the capital as says: "You really screwed up. In my village *every single one* of the T sampled children had a height-for-age Z score, Y_{19t}, less than negative -1. My villages' mean was $\bar{Y}_{19} = -5/4$, but because your goofy econometrician decided to shrink all the village means toward the country-wide mean of $\mu = 0$ (with $\hat{\lambda}^* = 1/5$), they reported $\hat{\theta}_{19} = -1$ to you. As a result I have a bunch of hungry kids not getting the help they need. Your estimator is biased!" Write a response to this Mayor's concern.

Part II: Quantile regression

[1] 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

$$Q_{U|X}\left(1/2 |\, X\right) = 0.$$

Let a and b be a candidate values for 'the truth' (i.e., α_0 and β_0). Let $u_{1/2}^{1}(a, b)$ be the median of $U(a, b) = Y - a - bX$ given $X = 1$. Let $u_{1/2}^{0}(a, b)$ 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] 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 median of $U(\alpha_0, \beta_0)$ given, respectively, $X = 1$ and $X = 0$?

[c] Assume that $a = \alpha_0$ and $b = \beta_0$. Let $j/(N_1 + 1) < 1/2 \leq (j+1)/(N_1+1)$. 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] Let $N_1 = 3$ 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.

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

Part III: Discrete hazard analysis

[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 world-

wide. 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

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

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

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

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)	$N_{(k)}$	$N^{\text{d}}_{(k)}$	$N^{\text{c}}_{(k)}$	$\lambda(y)$	$S(y)$
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 $\theta = \left(\gamma_0, \gamma_1, \beta\right)'$. 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.

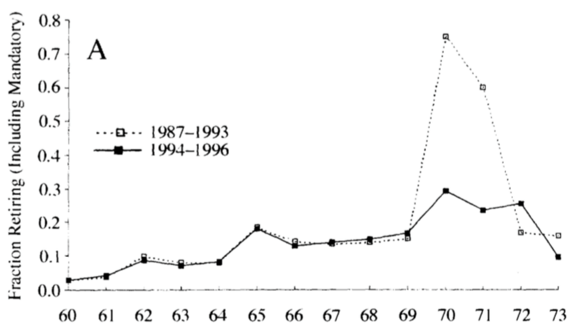
[2] Prior to 1994 colleges and universities in the United States were exempt from laws prohibiting mandatory retirement, consequently many institutions forced faculty to retire at age 70. After 1994 mandatory retirement rules were prohibited by Congress. Ashenfelter and Card (*AER*, 2002) study the effects 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, $D = \mathbf{1}(Y \leq C)$ be a censoring indicator, and $Z = \min(Y, C)$ 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

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

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 $S(y|X) = \Pr(Y > y|X)$. Does $\Pr(Z > y|X) = S(y|X)$? If not, does $\Pr(Z > y|X) > S(y|X)$ or $\Pr(Z > y|X) < S(y|X)$? Why? Describe a method for constructing an estimate of $S(y|X)$. 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 $S(y|X = 1)$ and $S(y|X = 0)$ 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

Age	Number of observations	Percentage post-1994	Average retirement rate: 1987–1993	Average retirement rate: 1994–1996	Change in retirement rate: Unadjusted	Change in retirement rate: 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.

Part IV: Regression redux

[1] Let Y equal tons of banana's harvested in a given season for a randomly sampled Honduran banana farm. 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 farm 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 farms (e.g., due to transportation costs, labor market segmentation etc.). We will treat D as a fixed factor; A captures sources of farm-level differences in farm productivity due to unobserved differences in, for example, soil quality and managerial capacity. Farm 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

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

[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

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

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] Briefly discuss the content and plausibility of the restriction

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

[d] 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.

[e] 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.

[f] Assume that all farms 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?

[g] Using the restriction in part [c], construct an instrumental variables estimate of α_0.

[2] Consider a population of firms. A firm with capital k and labor l produces output

$$y(k, l; A) = A k^{\alpha} l^{\beta}. \tag{7}$$

Here A is firm-specific, capturing heterogeneity in the efficiency with which different firms are able to transform capital and labor into output. Let W denote the wage rate, and R the rental price of a unit of capital, faced by the firm. We assume that a firm seeking to produce output y, while facing input prices r and w, does so in a *cost-minimizing* way; choosing $K(y,r,w;A)$ and $L(y,r,w;A)$ to solve the constrained minimization problem

$$\min_{k,l} rk + wl + \lambda\left[y - Ak^{\alpha}l^{\beta}\right]. \tag{8}$$

[a] The *derived demand* for capital and labor equals the amount of each input a firm would choose when seeking to produce y while facing input prices r and w. *Show* these derived demand schedules, under (8), equal:

$$K(y,r,w;A) = \alpha\left(\frac{y}{A}\right)^{\frac{1}{\eta}}\left(\frac{w}{r}\right)^{\frac{\beta}{\eta}}\left[\alpha^{\alpha}\beta^{\beta}\right]^{-\frac{1}{\eta}} \tag{9}$$

$$L(y,r,w;A) = \beta\left(\frac{y}{A}\right)^{\frac{1}{\eta}}\left(\frac{r}{w}\right)^{\frac{\alpha}{\eta}}\left[\alpha^{\alpha}\beta^{\beta}\right]^{-\frac{1}{\eta}}, \tag{10}$$

where $\eta = \alpha + \beta$. Briefly interpret the two demand schedules [3 to 4 sentences]. Why is η called the returns-to-scale parameter?

[b] The cost function equals

$$c(y,r,w;A) = rK(y,r,w;A) + wL(y,r,w;A). \tag{11}$$

Show that under (7) that this function equals

$$c(y,r,w;A) = \eta\left(\frac{y}{A}\right)^{\frac{1}{\eta}} r^{\frac{\alpha}{\eta}} w^{\frac{\beta}{\eta}}\left[\alpha^{\alpha}\beta^{\beta}\right]^{-\frac{1}{\eta}}. \tag{12}$$

Show that (12) is homogenous of degree one in input prices. Provide an economic explanation for this [4 to 6 sentences]? Show that the *cost shares* of capital and labor are, respectively α/η and β/η.

[c] Let $i = 1, \ldots, N$ index a random sample of firms. For each firm we observe output, Y_i, the input prices W_i and R_i, and total costs C_i. We do not observe the firm-specific productivity parameter, A_i. We assume that firm behavior is governed by (7), (9), (10) and (12). Imposing the restriction that (12) is homogenous of degree one in input prices and taking logs yields

$$\ln C_i - \ln W_i = \kappa_c + \frac{1}{\eta}\ln Y_i + \frac{\alpha}{\eta}\left[\ln R_i - \ln W_i\right] - \frac{1}{\eta}\left(\ln A_i - \mathbb{E}\left[\ln A_i\right]\right)$$

for $\kappa_c = \ln\left[\eta\left[\alpha^{\alpha}\beta^{\beta}\right]^{-\frac{1}{\eta}}\right] - \frac{1}{\eta}\mathbb{E}\left[\ln A_i\right]$. Consider the linear regression of log costs onto the (logs of) output and rents minus wages:

$$\mathbb{E}^*\left[\ln C_i - \ln W_i \middle| \ln Y_i, \ln R_i - \ln W_i\right] = k_0 + c_0 \ln Y_i + a_0\left[\ln R_i - \ln W_i\right]. \tag{13}$$

Is it likely that $c_0 = 1/\eta$ and $a_0 = \alpha/\eta$? Explain (mathematical calculations are not required; 6 to 8 sentences). So far our analysis has been silent regarding how firms choose their output level. In answering this question it might be helpful to consider two cases. In one case firms *do not* choose their output level (consider an electric utility that must meet demand at regulated prices). In the second case firms *do* choose their output level.