Second Midterm

Ec240a – Second Half, Fall 2019

Please read each question carefully. Start each question on a new bluebook page (or sheet of paper). 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] **[25 Points]** Let $X \in \{0, 1, 2\}$ and $Y \in \{0, 1, 2\}$. The probability of the event $X = x$ and $Y = y$ for all possible combinations of x and y is given in the following table:

X\Y	0	1	2
0	$\frac{15}{100}$	$\frac{10}{100}$	$\frac{5}{100}$
1	$\frac{10}{100}$	$\frac{20}{100}$	$\frac{10}{100}$
2	$\frac{5}{100}$	$\frac{10}{100}$	$\frac{15}{100}$

[a] **[5 Points]** Calculate $\mathbb{E}[Y]$ and $\mathbb{E}[Y|\, X = 1]$. Are X and Y independent?

[b] **[5 Points]** Calculate $\mathbb{E}[X]$, $\mathbb{E}[X^2]$ and hence $\mathbb{V}(X)$.

[c] **[5 Points]** Calculate $\mathbb{C}(X, Y)$ and also the coefficient on X in $\mathbb{E}^*[Y|\, X]$.

[d] **[5 Points]** Calculate the intercept of $\mathbb{E}^*[Y|\, X]$.

[e] **[5 Points]** Repeat [a] to [d] above for the following joint distribution

X\Y	0	1	2
0	$\frac{1}{6}$	0	0
1	0	$\frac{2}{3}$	0
2	0	0	$\frac{1}{6}$

[3] **[40 Points]** This question is about the Rambly Shambly Hex Bolt Corporation. Let Y_t be the number of hex bolts produced by Rambly Shambly Hex in year t, M_t tons of steel used in production, K_t total factory capital stock, and L_t total person-hours worked. We assume that

$$Y_t = A_t M_t^{\alpha} K_t^{\beta} L_t^{\gamma}.$$

[a] **[10 Points]** Rambly Shambly Hex is owned by an eccentric billionaire who chooses M_t, K_t and L_t each year randomly using an eternally unchanging roulette-wheel-like-device (i.e., inputs are chosen independently of each other and independently of A_t). Further assume

that the distribution of A_t is i.i.d. over time. Show that under this input choice mechanism that

$$\mathbb{E}^*\left[\ln Y_t|\ln M_t, \ln K_t, \ln L_t\right] = \lambda + \alpha \ln M_t + \beta \ln K_t + \gamma \ln L_t$$

with $\lambda = \mathbb{E}\left[\ln A_t\right]$. Is this same result likely to hold if Rambly Shambly instead chose input levels to maximize profits? Why or why not?

[b] **[7 Points]** Further show that under the completely random input choice scheme described above that:

$$\mathbb{E}^*\left[\ln Y_t|\ln M_t, \ln K_t, \ln L_t\right] = \mathbb{E}^*\left[\ln Y_t|\ln M_t\right] + \mathbb{E}^*\left[\ln Y_t|\ln K_t\right] + \mathbb{E}^*\left[\ln Y_t|\ln L_t\right] - 2\mathbb{E}\left[\ln Y_t\right].$$

[c] **[8 Points]** Let $X_t = (\ln M_t, \ln K_t, \ln L_t)'$ and $\sigma^2 = \mathbb{V}(\ln A_t)$. Argue that under the completely random input choice scheme:

$$\sqrt{T}\left(\hat{\theta} - \theta_0\right) \xrightarrow{D} N\left(0, \Lambda\right),$$

for $\theta = (\alpha, \beta, \gamma)'$, $\Lambda = \sigma^2 \mathbb{V}\left(X_t\right)^{-1}$ and $\hat{\theta}$ estimated by the OLS fit of $\ln Y_t$ onto a constant and X_t for $t = 1, \ldots, T$. What are the values of the off-diagonal elements of $\mathbb{V}\left(X_t\right)$?

[d] **[15 Points]** For $T = 3,859$, an OLS fit gives

$$\hat{\theta} = \begin{pmatrix} 0.32 \\ 0.36 \\ 0.42 \end{pmatrix}, \quad \hat{\Lambda} = \begin{pmatrix} 5 & 1/200 & 2/1000 \\ 1/200 & 3 & 3/1000 \\ 2/1000 & 3/1000 & \frac{198}{100} \end{pmatrix}.$$

Construct a Wald Statistic (carefully explaining each step in the construction) for the null hypothesis of constant returns to scale (i.e., $H_0 : \alpha + \beta + \gamma = 1$). What is the appropriate reference distribution and critical value for a two-sided test with size $\alpha = 0.05$? Do you reject the null?

[4] **[30 Points]** Let $m(Z) = \mathbb{E}\left[X|Z\right]$ and consider the linear regression

$$\mathbb{E}^*\left[Y|X, m(Z), A\right] = \alpha_0 + \beta_0 X + \gamma_0 m(Z) + A.$$

[a] **[10 Points]** Show that

$$\mathbb{E}^*\left[m(Z)|X\right] = \delta_0 + \xi_0 X$$

with

$$\begin{aligned}\delta_0 &= (1-\xi_0)\,\mathbb{E}[X] \\ \xi_0 &= \frac{\mathbb{V}(\mathbb{E}[X|Z])}{\mathbb{E}[\mathbb{V}(X|Z)]+\mathbb{V}(\mathbb{E}[X|Z])}.\end{aligned}$$

[b] **[5 Points]** Assume the population under consideration is working age adults who grew up in the San Francisco Bay Area. Let Y denote adult log income, let X denote the log income of one's parents as a child and let Z be a vector of dummy variables denoting an individual's neighborhood of residence as a child. Provide an interpretation of ξ_0 as a measure of residential stratification by income.

[c] **[7 Points]** Establish the notation $\rho = \operatorname{corr}(A, X)$, $\mu_A = \mathbb{E}[A]$, $\mu_X = \mathbb{E}[X]$, $\sigma_A^2 = \mathbb{V}(A)$ and $\sigma_X^2 = \mathbb{V}(X)$. Show that

$$\mathbb{E}^*[Y|X] = \alpha_0 + \gamma_0(1-\xi_0)\,\mu_X + \left(\mu_A - \rho\frac{\sigma_A}{\sigma_X}\mu_X\right) + \left\{\beta_0 + \gamma_0\xi_0 + \rho\frac{\sigma_A}{\sigma_X}\right\}X.$$

[d] **[8 Points]** Your research assistant computes an estimate of $\mathbb{E}^*[Y|X]$ using random sample from San Francisco. She computes a separate estimate using a random sample from New York City. Assume that there is more residential stratification by income in New York than in San Francisco. How would you expect the intercept and slope coefficients to differ across the two regression fits?