Econometrics I
Mid-Semester Exam
Winter 2022
Instructor: Gaurav Arora

Maximum Marks: 100

Instruction: This is an open-book, open-notes and open-internet exam. Use your time wisely. Institute plagiarism policy will apply *as-is*.

Q1.

(a) (40 points) Consider a random sample of size N denoted as $\{x_i, y_i\}_{i=1}^{100}$ where X and Y are random variables. We know three distinct methods for deriving data-driven measures of association between these two variables.

I. A regression of Y on X, i.e., $y_i = \beta_{0y} + \beta_{xy} x_i + u_i$ where u_i is the random error.

II. A regression of X on Y, i.e., $x_i = \beta_{0x} + \beta_{yx} y_i + e_i$ where e_i is the random error.

III. A correlation coefficient, i.e., r.

Show that $r = \pm\sqrt{\hat{\beta}_{xy,OLS} \times \hat{\beta}_{yx,OLS}}$. Given the regression line of X on Y: $3X - 2Y = 5$ and regression line of Y on X: $X - 4Y = 7$, please find the coefficient of correlation, r, between X and Y.

(b) (40 points)

Define a covariance measure derived from the data points of two arbitrary random variables L, M indexed by $i = 1, 2, \ldots, N$: $\text{cov}(l,m)_{i=1}^{N} = \dfrac{\sum_{i=1}^{N}(l_i - \bar{l})(m_i - \bar{m})}{N}$.

Evaluate the bias in OLS estimator $\hat{\beta}_{xy,OLS}$ when $\sigma_x^2 = 2$ and

I. $\text{cov}(x,u)_{i=1}^{100} = 0$.

II. $\text{cov}(x,u)_{i=1}^{100} = 2$

III. $\text{cov}(x,u)_{i=1}^{50} = 0$ and $\text{cov}(x,u)_{i=51}^{100} = -2$.

IV. $\text{cov}(x,u)_{i=1}^{50} = 0$, $\text{cov}(x,u)_{i=51}^{75} = -2$ and $\text{cov}(x,u)_{i=76}^{100} = 2$.

Q2. (20 points) Let's consider a shoe factory that uses web-based retail platforms for marketing and selling their products online. Inputs employed include internet connectivity (bytes of data) and electricity (kilowatt hours) in a manner that the number of pairs sold daily can be represented as

$$S_d = \eta I_d^{\alpha} E_d^{\beta} e^{\varepsilon_d} \qquad (1)$$

where

$S_d =$ total number of shoe pairs sold on day *d*,

I_d = total data bytes used on day *d*,

E_d = total kilowatt hours of electricity consumed on day *d*,

$\varepsilon_d \overset{iid}{\sim} N\left(0, \sigma^2\right)$, and

η, α and β are model parameters.

(a) Provide a convenient transformation of model (1) to estimate its parameters using the method of least squares? Please clearly describe the steps involved and also provide the estimates for *all* model parameters.

Now assume that you acquire 100 days of observations on the quantity sold, internet usage and electricity usage from a utility firm. Suppose the following information is provided to you by this firm:

$\tilde{\boldsymbol{S}}'\tilde{\boldsymbol{S}} = 290$ where $\tilde{\boldsymbol{S}}$ is a 100x1 vector of dependent variable containing daily values of $\tilde{S}_d = \log(S_d)$;

$\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{S}} = \begin{bmatrix} 166 \\ 180 \\ 123 \end{bmatrix}$, where $\tilde{\boldsymbol{X}}$ is a 100 x 3 vector of explanatory variables with first column as a vector of 1's and the second and third column as natural logarithms of daily internet and electricity usage. That is, $\tilde{X}_d = [1\ \tilde{I}_d\ \ \tilde{E}_d]$ where $\tilde{I}_d = \log(I_d)$; and $\tilde{E}_d = \log(E_d)$. Further, you are given that

$$\left(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\right)^{-1}=\begin{bmatrix} 0.06 & -0.05 & -0.002 \\ -0.05 & 0.1 & -0.07 \\ -0.002 & -0.07 & -0.12 \end{bmatrix}.$$

(b) Please provide an estimate of the parameters η, α and β given above information.

(c) How will you estimate $\hat{\sigma}^2$? Clearly describing in steps is a sufficient answer.