Advanced Econometrics
Midterm Exam, 2020
2 Hours and 15 Minutes

Problem 1. (10 Points) The OLS objective function discussed in class is

$$Q(a,b) = \sum_{i=1}^{n} (Y_i - a - bX_i)^2 .$$

Now consider a modification of it:

$$\widetilde{Q}(a,b) = \left[\sum_{i=1}^{n} (Y_i - a - bX_i)\right]^2 .$$

Let $\left(\widetilde{\alpha}, \widetilde{\beta}\right)$ be the minimizer of $\widetilde{Q}$. Show that $\widetilde{Q}\left(\widetilde{\alpha}, \widetilde{\beta}\right) = 0$. Hint: You do not need to derive the first order conditions.

Problem 2. (10 Points) Let Y and X be two random variables.

(i) Show that $\mathbb{E}(Y|X)$ and $Y - \mathbb{E}(Y|X)$ are uncorrelated. Hint: Use law of iterated expectations.

(ii) Show that $\text{Var}(Y) \geq \text{Var}(Y - \mathbb{E}(Y|X))$. Hint: Use (i).

Problem 3. (10 Points) Let $\boldsymbol{X}$ be the matrix collecting all the n observations on the k regressors. Let $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B}$, where $\boldsymbol{B}$ is a $k \times k$ non-singular matrix. Let $\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{e}}\right)$ denote the LS estimates and residuals from regression of $\boldsymbol{Y}$ on $\boldsymbol{X}$. Similarly, let $\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{e}}\right)$ denote these from regression of $\boldsymbol{Y}$ on $\boldsymbol{Z}$. Find the relationship between $\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{e}}\right)$ and $\left(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{e}}\right)$.

Problem 4. (20 Points) Consider a regression of Y_i against a constant and X_i. Let $\hat{\beta}_0$, $\hat{\beta}_1$, and s^2 denote the estimated intercept, estimated slope parameter, and estimator of the variance of errors from that regression. Let T denote the t-statistic for testing H_0 that the slope parameter is zero in that regression.
Now, let c_1 and c_2 be two constants ($c_2 \neq 0$). Define a new dependent variable and a new regressor as

$$\begin{aligned} Y_i^* &= c_1 Y_i, \\ X_i^* &= c_2 X_i. \end{aligned}$$

Let $\hat{\beta}_0^*$, $\hat{\beta}_1^*$, and s_*^2 denote the estimated intercept, estimated slope parameter, and estimator of the variance of errors from the regression of Y_i^* against a constant and X_i^*. Let T^* denote the t-statistic for testing H_0 that the slope parameter in the regression of Y_i^* against a constant and X_i^* is zero.

(i) Find an expression for $\hat{\beta}_1^*$ in terms of $\hat{\beta}_1, c_1$, and c_2.

(ii) Find an expression for $\hat{\beta}_0^*$ in terms of $\hat{\beta}_0$ and c_1.

(iii) Find an expression for s_*^2 in terms of s^2 and c_1.

(iv) What is the relationship between T and T^*?

Problem 5. (15 Points) Show that in a simple (one-regressor) regression model,

$$Y_i = \beta_0 + \beta_1 X_i + U_i, \; i = 1, \ldots, n,$$

the LS estimate for β_1 is

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n \left(X_i - \overline{X}\right) Y_i}{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2},$$

where $\overline{X} = n^{-1} \sum_{i=1}^n X_i$. Then assume (1) (X_i, Y_i), $i = 1, ..., n$ are independently and identically distributed (i.i.d.). (2) $E\left(U_i | X_i\right) = 0$, for $i = 1, ..., n$. (3) $E\left(U_i^2 | X_i\right) = \sigma^2$, for $i = 1, ..., n$, with some $\sigma > 0$. Show that

$$\text{Var}\left(\widehat{\beta}_1 | X_1, ..., X_n\right) = \frac{\sigma^2}{\sum_{i=1}^n \left(X_i - \overline{X}\right)^2}.$$

Problem 6. (15 Points) Consider a simple linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + U_i,$$

where Y_i is the dependent variable, X_i is the explanatory variable, β_0 and β_1 are unknown scalar parameters, $\{(X_i, U_i) : i = 1, \ldots, n\}$ are independent and identically distributed, $E\left(U_i | X_i\right) = 0$ and $E\left(U_i^2 | X_i\right) = \sigma^2$. Consider the following estimator of β_1: $\widetilde{\beta}_1 = \frac{Y_n - Y_1}{X_n - X_1}$, where (X_1, Y_1) and (X_n, Y_n) are the first and last observations in the data set respectively. Assume that

$$P\left(X_1 \neq X_n\right) = 1.$$

(i) Is $\widetilde{\beta}_1$ unbiased?

(ii) Find $Var\left(\widetilde{\beta}_1 | X_1, \ldots, X_n\right)$.

(iii) Let $\hat{\beta}_1$ be the OLS estimator of β_1. Show directly, without relying on the Gauss-Markov theorem, that the conditional variance of $\hat{\beta}_1$ is smaller than that of $\widetilde{\beta}_1$. Hints: First, show that $Var\left(\widetilde{\beta}_1 | X_1, \ldots, X_n\right) > Var\left(\hat{\beta}_1 | X_1, \ldots, X_n\right)$ if

$$\sum_{i=1}^n \left(X_i - \overline{X}\right)^2 - \frac{1}{2}\left(\left(X_n - \overline{X}\right) - \left(X_1 - \overline{X}\right)\right)^2 > 0, \tag{1}$$

and then prove that the inequality in (1) holds.

Problem 7. (10 Points) Let $\boldsymbol{U} \mid \boldsymbol{X} \sim N(0, \sigma^2 \boldsymbol{I}_n)$, where $\boldsymbol{X}$ is an $n \times k$ matrix of rank k, and $\sigma^2 > 0$ is an unknown constant.

(i) Describe the distribution of $\boldsymbol{U}'\boldsymbol{P}\boldsymbol{U}/\sigma^2$, where $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$.

(ii) Give a proof of your result in part (i). Show how the result is implied by the fact that $\boldsymbol{P}$ is symmetric and idempotent.

Problem 8. (10 Points) Suppose we observe a random sample $\{(Y_i, D_i)\}_{i=1}^n$, where Y_i is the dependent variable and D_i is a binary independent variable: for all $i = 1, 2, ..., n$, $D_i = 1$ or $D_i = 0$. Suppose we regress Y_i on D_i with an intercept. Show: the LS estimate of the slope is equal to the difference between the sample averages of the dependent variable of the two groups, observations with $D_i = 1$ and observations with $D_i = 0$. Hint: The sample average of Y of observations with $D_i = 1$ can be written as $\frac{\sum_{i=1}^n D_i Y_i}{\sum_{i=1}^n D_i}$. What is the sample average of Y of observations with $D_i = 0$? Also note: $D_i = D_i^2$.