Advanced Econometrics
Midterm Exam, 2021 (6 questions, 2 hours)

Problem 1. (15 points) Partition the matrix of regressors $\boldsymbol{X}$ as follows:

$$\boldsymbol{X} = [\boldsymbol{X}_1 \ \ \boldsymbol{X}_2].$$

$\boldsymbol{M}_1 = \boldsymbol{I}_n - \boldsymbol{X}_1 \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1} \boldsymbol{X}_1'$. Consider the regression model

$$\boldsymbol{Y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e}.$$

Let $\widehat{\boldsymbol{\beta}}_1$ and $\widehat{\boldsymbol{\beta}}_2$ denote the least squares estimates from running the regression:

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{pmatrix} = \underset{(\boldsymbol{b}_1,\boldsymbol{b}_2)\in\mathbb{R}^{k_1}\times\mathbb{R}^{k_2}}{\operatorname{argmin}} \left(\boldsymbol{Y} - \boldsymbol{X}_1\boldsymbol{b}_1 - \boldsymbol{X}_2\boldsymbol{b}_2\right)' \left(\boldsymbol{Y} - \boldsymbol{X}_1\boldsymbol{b}_1 - \boldsymbol{X}_2\boldsymbol{b}_2\right).$$

Denote $\widehat{\boldsymbol{e}} = \boldsymbol{Y} - \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 - \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2$. Consider the following regressions, all to be estimated by least squares:

(i) Regress $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Y}$ on $\boldsymbol{X}_2$. Let $\widetilde{\boldsymbol{\beta}}_2$ denote the least squares estimates.

(ii) Regress $\boldsymbol{Y}$ on $\boldsymbol{M}_1\boldsymbol{X}_2$. Let $\widetilde{\boldsymbol{\beta}}_2$ denote the least squares estimates.

(iii) Regress $\boldsymbol{P}_{\boldsymbol{X}}\boldsymbol{Y}$ on $\boldsymbol{M}_1\boldsymbol{X}_2$. Let $\widetilde{\boldsymbol{\beta}}_2$ denote the least squares estimates.

For which of the above regressions will the estimates be the same as $\widehat{\boldsymbol{\beta}}_2$? For which will the residuals be the same as $\widehat{\boldsymbol{e}}$?

Problem 2. (15 points) Consider the following model: $Y = \alpha + \beta X + U$, where $\mathbb{E}(U \mid X) = 0$, $\mathbb{E}\left(U^2 \mid X\right) = \sigma^2 > 0$ and the conditional distribution of U given X is $\mathrm{N}\left(0, \sigma^2\right)$.

(i) What is $\mathbb{E}(Y \mid X)$? What is the conditional distribution of Y given X?

(ii) In this question, use the following result: if the conditional distribution of Y given X is $\mathbb{E}\left(m(X), s(X)^2\right)$ ($s(X) > 0$), then the conditional distribution of $\frac{Y-m(X)}{s(X)}$ given X is $N(0,1)$ and

$$\Pr\left(\frac{Y - m(X)}{s(X)} \leq z \mid X\right) = \Phi(z),$$

where $\Phi(z)$ is the standard normal CDF. Here, the left hand side means the conditional probability of $\frac{Y-m(X)}{s(X)} \leq z$ given X. For a random variable Z, its τ-th quantile ($\tau \in (0,1)$) q_τ is defined by the equation: $\Pr(Z \leq q_\tau) = \tau$. Similarly, a function $q_\tau(X)$ is the τ-th quantile of the conditional distribution of Y given X if

$$\Pr(Y \leq q_\tau(X) \mid X) = \tau.$$

Find the expression of $q_\tau(X)$. **Hint**: Let z_τ denote the τ-th quantile of the standard normal distribution so that $\Phi(z_\tau) = \tau$.

(iii) Suppose that $\mathbb{E}\left(U^2 \mid X\right) = e^{2X}$ and the conditional distribution of U given X is $\mathrm{N}\left(0, e^{2X}\right)$. Find the expression of $q_\tau(X)$.

Problem 3. (34 points) Consider the simple (one-regressor) linear regression model without an intercept:

$$\boldsymbol{Y} = \beta \boldsymbol{X} + \boldsymbol{e},$$

where $\boldsymbol{Y}$, $\boldsymbol{X}$, and $\boldsymbol{e}$ are n-dimensional random vectors, and β is an unknown scalar parameter. Assume that $\mathbb{E}(\boldsymbol{e} \mid \boldsymbol{X}) = 0$ and $\mathbb{E}(\boldsymbol{e}\boldsymbol{e}' \mid \boldsymbol{X}) = \sigma^2 \boldsymbol{I}_n$.

(i) Show that the least squares estimator of β is

$$\widehat{\beta} = \frac{\boldsymbol{X}'\boldsymbol{Y}}{\boldsymbol{X}'\boldsymbol{X}}.$$

(ii) Define the fitted residuals $\widehat{\boldsymbol{e}} = \boldsymbol{Y} - \widehat{\beta}\boldsymbol{X}$. $\boldsymbol{Y} = (Y_1, ..., Y_n)'$, $\boldsymbol{X} = (X_1, ..., X_n)'$, $\boldsymbol{e} = (e_1, ..., e_n)'$, $\widehat{\boldsymbol{e}} = (\widehat{e}_1, ..., \widehat{e}_n)'$. For each of the following statements, explain if it is true or false:

(a) $\mathbb{E}(e_i X_i) = 0$ for all $i = 1, \ldots, n$.
(b) $\mathbb{E}e_i = 0$ for all $i = 1, \ldots, n$.
(c) $\sum_{i=1}^n \widehat{e}_i X_i = 0$.
(d) $\sum_{i=1}^n \widehat{e}_i = 0$.
(e) $\sum_{i=1}^n e_i X_i = 0$.
(f) $\sum_{i=1}^n e_i = 0$.

(iii) Find $\text{Var}\left(\widehat{\beta} \mid \boldsymbol{X}\right)$.

(iv) Consider the following estimator of β:

$$\widetilde{\beta} = \frac{\bar{Y}}{\bar{X}},$$

where

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i \text{ and } \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i.$$

Is $\widetilde{\beta}$ unbiased?

(v) Find $\text{Var}\left(\widetilde{\beta} \mid \boldsymbol{X}\right)$.

Problem 4. (8 points) Consider a partitioned linear regression model $\boldsymbol{Y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{e}$, and assume that the assumptions of the classical normal linear regression hold. Let $\widehat{\boldsymbol{\beta}}_1$ be the OLS estimator of $\boldsymbol{\beta}_1$. Show that $\text{Var}(\widehat{\boldsymbol{\beta}}_1 \mid \boldsymbol{X}_1, \boldsymbol{X}_2) = \sigma^2(\boldsymbol{X}_1'\boldsymbol{M}_2\boldsymbol{X}_1)^{-1}$, where $\boldsymbol{M}_2 = \boldsymbol{I}_n - \boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'$.

Problem 5. (20 points) We are studying the factors behind the Body Mass Index (BMI), defined as the weight (in kilograms) divided by the squared of height (in meters). The number of observations is 233239. For a random sample we have estimated the following model (**Model 1**, **Output 1**):

$$\widehat{BMI} = \widehat{\beta}_0 + \widehat{\beta}_1 * drinks + \widehat{\beta}_2 * drinks^2 + \widehat{\beta}_3 * female$$

where

$drinks$	=	number of days during the last year in which the individual has drunk 5 or more glasses of alcohol.
$drinks^2$	=	squared of drinks.
$female$	=	is a dummy variable that takes a value one for women and zero otherwise.

Alternatively we have estimated the following model (**Model 2**, **Output 2**):

$$\widehat{BMI} = \widehat{\beta}_0 + \widehat{\beta}_1 * drinks + \widehat{\beta}_2 drinks^2 + \widehat{\beta}_3 * female + \widehat{\beta}_4 * drinks * female + \widehat{\beta}_5 * drinks^2 * female.$$

Useful the critical value $F_{2,N,1-\alpha} = 3.00$ when $N > 200$.

(i) Using Model 1, What is the marginal effect of drinks on expected BMI? Is this effect linear or constant? Perform a t test to answer this question.

(ii) Using Model 1, What is the predicted difference in expected BMI between a man with $drinks = 2$ and a woman with $drinks = 6$?

(iii) Using Model 2 as benchmark (unrestricted model), explain and test whether the effect of a marginal change in drinks on BMI is constant for men.

(iv) Using Model 2 as benchmark (unrestricted model), explain and test whether the effect of a marginal change in drinks on BMI is constant for an individual whatever the gender. **Hint**: use Output 4.

Table 1: OUTPUT 1

Dependent variable			*BMI*	
	coefficient	**S.E**	t	p-**value**
constant	26.8065	0.0182106	1472.0235	0.0000
drinks	–0.00953349	0.00484486	-1.9678	0.0491
$drinks^2$	0.000102875	4.98842e–005	2.0623	0.0392
female	–1.14183	0.020171		
SSR			5196636	
R^2			0.014139	

Table 2: OUTPUT 2

Dependent variable			*BMI*	
	coefficient	**S.E**	t	p-**value**
constant	26.6876	0.0197415	1351.8542	0.0000
drinks	0.0424549	0.00586037	7.2444	0.0000
$drinks^2$	0.000419308	6.03490e–005	6.9481	0.0000
female	0.875380	0.0264042	–33.1531	0.0000
$drinks \times female$	–0.163936	0.0104052	–15.7552	0.0000
$drinks^2 \times female$	0.00165051	0.000107120		
SSR			5191109	
R^2			0.015188	

Table 3: OUTPUT 4

Dependent variable			*BMI*	
	coefficient	**S.E**	t	p-**value**
constant	26.1650	0.0102729	2546.9867	0.0000
drinks	0.0214939	0.00139848	15.3695	0.0491
$drinks \times female$	–0.0445572	0.00232373	–19.1749	0.0000
female	–1.113876	0.0205		
SSR			5262194	
R^2			0.001702	

Problem 6. (8 points) Suppose that we observe a random sample $X_1, ..., X_n$ where $X_1, ..., X_n$ are independent $N\left(\mu, \sigma^2\right)$ random variables. σ^2 is known. Consider the test statistic ($\overline{X} = n^{-1}\sum_{i=1}^{n} X_i$)

$$T = \frac{\overline{X}}{\sigma/\sqrt{n}}$$

for the two-sided hypothesis test $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$. H_0 is rejected if $|T| > z_{1-\alpha/2}$ where $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$-th quantile of $N(0, 1)$. Give the expression of the power function as a function of μ. Use Φ to denote the standard normal CDF.