Introductory Econometrics Midterm Exam

Problem 1. (15 points) Consider an equation to explain salaries of CEOs in terms of annual sales, return on equity (roe, in percentage form), and return on the firm's stock(ros, in percentage form):

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 roe + \beta_3 ros + u.$$

1. In terms of the model parameters, state the null hypothesis that, after controlling for $sales$ and roe, ros has no effect on CEO salary. State the alternative that better stock market performance increases a CEO's salary.

2. The following equation was obtained by OLS:

$$\begin{aligned}\log(\widehat{salary}) &= 4.32 + 0.280 \log(sales) + 0.0174 roe + 0.00024 ros \\ &\quad (0.32) \qquad (0.035) \qquad (0.0041) \qquad (0.00054) \\ &\quad n = 209,\ R^2 = 0.283.\end{aligned}$$

   By what percentage is $salary$ predicted to increase if ros increase by 50 points? Does ros have a practically large effect on $salary$?

3. Test the null hypothesis that ros has no effect on $salary$ against the alternative that ros has a positive effect. Carry out the test at the 5% significance level. Use one of the critical values: 1.64 or -1.64.

Solution.

1. $H_0 : \beta_3 = 0$. $H_1 : \beta_3 > 0$.

2. The proportionate effect on $\widehat{salary}$ is $0.00024 \times 50 = 0.012$. To obtain the percentage effect, we multiply this by 100: 1.2%. Therefore, a 50 point increase in ros is predicted to increase salary by only 1.2%. Practically speaking, this is a very small effect for such a large change in ros.

3. The t statistic on ros is $0.00024/0.00054 \approx 0.44$, which is well below the critical value 1.64. Therefore, we fail to reject H_0 at the 10% significance level.

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

(a) (5 points) What is $E(Y|X)$? What is the conditional distribution of Y given X?

(b) (5 points) In this question, use the following result: if the conditional distribution of Y given X is $N\left(m(X), s(X)^2\right)$ ($s(X) > 0$), then the conditional distribution of $\frac{Y-m(X)}{s(X)}$ given X is $N(0,1)$ and

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

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

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

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

(c) (5 points) Suppose that $E\left(U^2|X\right) = e^{2X}$ and the conditional distribution of U given X is $N\left(0, e^{2X}\right)$. Find the expression of $q_\tau(X)$.

Problem 3. (15 points) Consider the following regression model

$$\begin{aligned}
Y_i &= \beta X_i + U_i,\ i = 1, \ldots, n; \\
E(U_i|X_1, \ldots, X_n) &= 0; \\
E\left(U_i^2|X_1, \ldots, X_n\right) &= \sigma^2; \\
E(U_iU_j|X_1, \ldots, X_n) &= 0 \text{ for all } i \neq j.
\end{aligned}$$

Assume that $\bar{X} \neq 0$ and consider the following estimator of β:

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

(a) (5 points) Show that $\tilde{\beta}$ is unbiased.

(b) (5 points) Show that

$$Var\left(\tilde{\beta}|X_1, \ldots, X_n\right) = \frac{\sigma^2}{n\left(\bar{X}\right)^2}.$$

(c) (5 points) Let

$$\hat{\beta} = \frac{\sum_{i=1}^n X_iY_i}{\sum_{i=1}^n X_i^2},$$

and recall that

$$Var\left(\hat{\beta}|X_1, \ldots, X_n\right) = \frac{\sigma^2}{\sum_{i=1}^n X_i^2}.$$

Without using the Gauss-Markov Theorem, show that

$$Var\left(\hat{\beta}|X_1, \ldots, X_n\right) \leq Var\left(\tilde{\beta}|X_1, \ldots, X_n\right).$$

Hint: What is the relationship between $\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$, $\sum_{i=1}^{n}X_i^2$, and $n\left(\bar{X}\right)^2$?

Problem 4. (10 points) Find A-C in the Stata output below. The number of observations is 29. Justify your answer. According to the results, is x significant? The significance level is 5%. Explain.

```
------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |          A    .373891       B    0.020     .2025939           C
       _cons |   1.261567   .1814382     6.95   0.000     .8892861    1.633847
------------------------------------------------------------------------------
```

Problem 5. (20 points) Consider the following estimated regression model:

$$Y_i=\hat{\beta}_0+\hat{\beta}_1X_{1i}+\hat{\beta}_2X_{2i}+\hat{U}_i,$$

where $\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2$ are the OLS estimators and $\hat{U}_i$'s are the OLS residuals, i.e. $\hat{\beta}_0,\hat{\beta}_1$, and $\hat{\beta}_2$ are such that for $\hat{U}_i=Y_i-\hat{\beta}_0-\hat{\beta}_1X_{1i}-\hat{\beta}_2X_{2i}$,

$$\sum_{i=1}^{n}\hat{U}_i=\sum_{i=1}^{n}\hat{U}_iX_{1i}=\sum_{i=1}^{n}\hat{U}_iX_{2i}=0.$$

Let $\tilde{X}_{2i}$ be the OLS residuals from the regression of X_{2i} against a constant and X_{1i}:

$$\tilde{X}_{2i}=X_{2i}-\hat{\gamma}_0-\hat{\gamma}_1X_{1i},$$

where $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are such that

$$\sum_{i=1}^{n}\tilde{X}_{2i}=\sum_{i=1}^{n}\tilde{X}_{2i}X_{1i}=0.$$

Let $\tilde{Y}_i$ be the OLS residuals from the regression of Y_i against a constant and X_{1i}:

$$\tilde{Y}_i=Y_i-\hat{\delta}_0-\hat{\delta}_1X_{1i},$$

where $\hat{\delta}_0$ and $\hat{\delta}_1$ are such that

$$\sum_{i=1}^{n}\tilde{Y}_i=\sum_{i=1}^{n}\tilde{Y}_iX_{1i}=0.$$

(a) (5 points) In class, we showed that $\hat{\beta}_2$ can be obtained from a regression of Y_i against $\tilde{X}_{2i}$ without a constant:

$$\hat{\beta}_2=\frac{\sum_{i=1}^{n}\tilde{X}_{2i}Y_i}{\sum_{i=1}^{n}\tilde{X}_{2i}^2}. \tag{1}$$

Prove that the residuals $\bar{U}_i=Y_i-\tilde{X}_{2i}\hat{\beta}_2$ from this regression are not necessarily the same as $\hat{U}_i$'s. Hint: consider $\sum_{i=1}^{n}\bar{U}_i$ and $\sum_{i=1}^{n}\hat{U}_i$.

(b) (5 points) Show that $\hat{\beta}_2$ can be obtained from a regression of $\tilde{Y}_i$ against $\tilde{X}_{2i}$ without a constant:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n \tilde{X}_{2i}\tilde{Y}_i}{\sum_{i=1}^n \tilde{X}_{2i}^2}.$$

Hint: Use (1).

(c) (10 points) Let $\tilde{U}_i = \tilde{Y}_i - \tilde{X}_{2i}\hat{\beta}_2$ denote the residuals from the regression in Part (b). Prove that $\tilde{U}_i = \hat{U}_i$ for all $i = 1, ..., n$. Hint: Use

$$\hat{U}_i - \tilde{U}_i = \left(\hat{\delta}_0 - \hat{\beta}_0\right) + \left(\hat{\delta}_1 - \hat{\beta}_1\right) X_{1i}. \tag{2}$$

Show that

$$\sum_{i=1}^n \tilde{U}_i^2 = \sum_{i=1}^n \hat{U}_i^2 = \sum_{i=1}^n \tilde{U}_i\hat{U}_i$$

and $\sum_{i=1}^n \left(\tilde{U}_i - \hat{U}_i\right)^2 = 0$.

Solution.

(a) We have $\sum_{i=1}^n \hat{U}_i = 0$. But $\sum_{i=1}^n \bar{U}_i = \sum_{i=1}^n Y_i - \hat{\beta}_2 \sum_{i=1}^n \tilde{X}_{2i} = \sum_{i=1}^n Y_i$. So $\sum_{i=1}^n \bar{U}_i \neq 0$ unless $\sum_{i=1}^n Y_i = 0$. Therefore, $\bar{U}_i \neq \hat{U}_i$, in general.

(b)

$$\begin{aligned}\hat{\beta}_2 = \frac{\sum_{i=1}^n \tilde{X}_{2i}Y_i}{\sum_{i=1}^n \tilde{X}_{2i}^2} &= \frac{\sum_{i=1}^n \tilde{X}_{2i}\left(\hat{\delta}_0 + \hat{\delta}_1 X_{1i} + \tilde{Y}_i\right)}{\sum_{i=1}^n \tilde{X}_{2i}^2} \\ &= \frac{\hat{\delta}_0 \sum_{i=1}^n \tilde{X}_{2i} + \hat{\delta}_1 \sum_{i=1}^n \tilde{X}_{2i}X_{1i} + \sum_{i=1}^n \tilde{X}_{2i}\tilde{Y}_i}{\sum_{i=1}^n \tilde{X}_{2i}^2} = \frac{0 + 0 + \sum_{i=1}^n \tilde{X}_{2i}\tilde{Y}_i}{\sum_{i=1}^n \tilde{X}_{2i}^2}.\end{aligned}$$

(c) Note that $\sum_{i=1}^n \left(\tilde{U}_i - \hat{U}_i\right)^2 = 0$ if and only if $\tilde{U}_i = \hat{U}_i$ for all $i = 1, ..., n$. We have

$$\sum_{i=1}^n \left(\tilde{U}_i - \hat{U}_i\right)^2 = \sum_{i=1}^n \tilde{U}_i^2 + \sum_{i=1}^n \hat{U}_i^2 - 2\sum_{i=1}^n \tilde{U}_i\hat{U}_i,$$

$$\sum_{i=1}^n \tilde{U}_i^2 = \sum_{i=1}^n \tilde{U}_i\left(\tilde{U}_i - \hat{U}_i + \hat{U}_i\right) = \sum_{i=1}^n \tilde{U}_i\left(\tilde{U}_i - \hat{U}_i\right) + \sum_{i=1}^n \tilde{U}_i\hat{U}_i$$

and

$$\sum_{i=1}^n \hat{U}_i^2 = \sum_{i=1}^n \hat{U}_i\left(\hat{U}_i - \tilde{U}_i + \tilde{U}_i\right) = \sum_{i=1}^n \hat{U}_i\left(\hat{U}_i - \tilde{U}_i\right) + \sum_{i=1}^n \hat{U}_i\tilde{U}_i.$$

Now, by (2),

$$\sum_{i=1}^n \tilde{U}_i\left(\tilde{U}_i - \hat{U}_i\right) = -\left(\hat{\delta}_0 - \hat{\beta}_0\right)\sum_{i=1}^n \tilde{U}_i - \left(\hat{\delta}_1 - \hat{\beta}_1\right)\sum_{i=1}^n \tilde{U}_i X_{1i}$$

$$= -\left(\hat{\delta}_0 - \hat{\beta}_0\right) \sum_{i=1}^{n} \left(\tilde{Y}_i - \tilde{X}_{2i}\hat{\beta}_2\right) - \left(\hat{\delta}_1 - \hat{\beta}_1\right) \sum_{i=1}^{n} \left(\tilde{Y}_i - \tilde{X}_{2i}\hat{\beta}_2\right) X_{1i}$$

$$= -\left(\hat{\delta}_0 - \hat{\beta}_0\right) \left\{\sum_{i=1}^{n} \tilde{Y}_i - \hat{\beta}_2 \sum_{i=1}^{n} \tilde{X}_{2i}\right\} - \left(\hat{\delta}_1 - \hat{\beta}_1\right) \left\{\sum_{i=1}^{n} \tilde{Y}_i X_{1i} - \hat{\beta}_2 \sum_{i=1}^{n} \tilde{X}_{2i} X_{1i}\right\} = 0$$

since $\sum_{i=1}^{n} \tilde{Y}_i = \sum_{i=1}^{n} \tilde{X}_{2i} = \sum_{i=1}^{n} \tilde{Y}_i X_{1i} = \sum_{i=1}^{n} \tilde{X}_{2i} X_{1i} = 0$, and

$$\sum_{i=1}^{n} \hat{U}_i \left(\hat{U}_i - \tilde{U}_i\right) = \sum_{i=1}^{n} \hat{U}_i \left\{\left(\hat{\delta}_0 - \hat{\beta}_0\right) + \left(\hat{\delta}_1 - \hat{\beta}_1\right) X_{1i}\right\}$$

$$= \left(\hat{\delta}_0 - \hat{\beta}_0\right) \sum_{i=1}^{n} \hat{U}_i + \left(\hat{\delta}_1 - \hat{\beta}_1\right) \sum_{i=1}^{n} \hat{U}_i X_{1i} = 0$$

since $\sum_{i=1}^{n} \hat{U}_i = \sum_{i=1}^{n} \hat{U}_i X_{1i} = 0$. Therefore,

$$\sum_{i=1}^{n} \tilde{U}_i^2 = \sum_{i=1}^{n} \hat{U}_i^2 = \sum_{i=1}^{n} \tilde{U}_i \hat{U}_i.$$

Problem 6. (25 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

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

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 $N > 200$.

1. (6 points) 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.

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

3. (6 points) 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.

4. (7 points) 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	

Q1: (a). $E[Y|X] = \alpha + \beta X$

$$Y|X \sim N(\alpha + \beta X, \sigma^2)$$

$\Phi(z_\tau) = \tau$,

z_τ is the τ-th quantile of $N(0,1)$

(b). $$P\left(\frac{Y-(\alpha+\beta X)}{\sigma} \leq z_\tau \middle| X\right) = \tau$$

$$\Rightarrow P(Y \leq \sigma \cdot z_\tau + \alpha + \beta X | X) = \tau$$

$$\Rightarrow q_\tau(X) = \sigma z_\tau + \alpha + \beta X$$

(c). $$Y|X \sim N(\alpha + \beta X, e^{2X})$$

$$P\left(\frac{Y-(\alpha+\beta X)}{e^X} \leq z_\tau \middle| X\right) = \tau$$

$$\Rightarrow P(Y \leq z_\tau e^X + \alpha + \beta X \mid X) = \tau$$

$$\Rightarrow q_\tau(X) = z_\tau e^X + \alpha + \beta X$$

Q2: (a). $$\tilde{\beta} = \frac{\bar{Y}}{\bar{X}} = \frac{1}{\bar{X}} \cdot \frac{1}{n}\sum_{i=1}^{n}(\beta X_i + u_i)$$

$$= \frac{1}{\bar{X}} \cdot (\beta \bar{X} + \bar{u})$$

$$= \beta + \frac{\bar{u}}{\bar{X}}$$

$$E[\tilde{\beta} \mid X_1 \cdots X_n] = \beta + \frac{1}{\bar{X}} E[\bar{u} \mid X_1, \cdots, X_n]$$

$$= \beta + \frac{1}{\bar{X}} \cdot \frac{1}{n} \sum_{i=1}^{n} E[u_i \mid X_1, \cdots, X_n]$$

$$= \beta + \frac{1}{\bar{X}} \cdot 0$$

$$= \beta$$

By LIE, $E[\tilde{\beta}] = E[E[\tilde{\beta} \mid X_1 \cdots X_n]]$

$$= E[\beta]$$

$$= \beta$$

$\Rightarrow \tilde{\beta}$ is unbiased.

(b) $\mathrm{Var}(\tilde{\beta} \mid X_1 \cdots X_n)$

$$= E[(\tilde{\beta} - E[\tilde{\beta} \mid X_1 \cdots X_n])^2 \mid X_1 \cdots X_n]$$

$$= E[(\tilde{\beta} - \beta)^2 \mid X_1 \cdots X_n]$$

$$= E\left[\left(\frac{\bar{u}}{\bar{X}}\right)^2 \middle| X_1 \cdots X_n\right]$$

$$= \frac{1}{\bar{X}^2} E\left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} u_i u_j \middle| X_1 \cdots X_n\right]$$

$$= \frac{1}{n^2}\frac{1}{\bar{X}^2}\left\{E\left[\sum_{i=1}^{n} u_i^2 \middle| X_1 \cdots X_n\right]\right.$$

$$\left. + E\left[\sum_{i=1}^{n}\sum_{j\neq i} u_i u_j \middle| X_1 \cdots X_n\right]\right\}$$

$$= \frac{1}{n^2}\frac{1}{\bar{X}^2}\left(n\sigma^2 + 0\right),$$ since

$$E[u_i^2 | X_1 \cdots X_n] = \sigma^2 \quad \forall i$$

$$E[u_i u_j | X_1 \cdots X_n] = 0 \quad \forall i \neq j$$

$$= \frac{\sigma^2}{n(\bar{X})^2}$$

(c). $\sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 \geq 0$

$$\Longrightarrow \sum_{i=1}^{n} X_i^2 \geq n\bar{X}^2$$

Q3 :

A : $0.2026 = \hat{\beta}_1 - t_{27, 0.975} \times \text{standard error}$

$= \hat{\beta}_1 - 2.052 \times 0.3739$

$\Rightarrow \hat{\beta}_1 = 0.9698$

B : $t = \dfrac{\hat{\beta}_1}{\text{standard error}} = \dfrac{0.9698}{0.3739} \approx 2.5939$

C : $\hat{\beta}_1 + t_{27, 0.975} \times \text{standard error}$

$= 0.9698 + 2.052 \times 0.3739$

≈ 1.7371

p-value $= 0.02 < 5\% \Rightarrow x$ is significant.

Question 3.

(a). marginal effect $= \beta_1 + 2\beta_2 * drinks$

(7 points) //

$\frac{\partial \widehat{BMI}}{\partial drinks}$

(or $\hat{\beta}_1 + 2\hat{\beta}_2 * drinks$)

$H_0: \beta_2 = 0 \quad H_1: \beta_2 \neq 0$

t-test: from Table 1: OUTPUT 1,

$t = \underbrace{2.0623}_{>1.96}$, p-value $= 0.0392 < 0.05$

$\Rightarrow$ reject H_0.

$\Rightarrow$ the marginal effect is not constant

(b) (7 points)

$\widehat{E(BMI \mid female=0, drinks=2)}$

$= \hat{\beta}_0 + \hat{\beta}_1 \times 2 + \hat{\beta}_2 \times 4$

$\widehat{E(BMI \mid female=1, drinks=6)}$

$= \hat{\beta}_0 + \hat{\beta}_1 \times 6 + \hat{\beta}_2 \times 6^2 + \hat{\beta}_3$

predicted difference $= \hat{\beta}_1 (2-6) + \hat{\beta}_2 (4-36) - \hat{\beta}_3$

$= 0.0095 \times 4 - 0.0001 \times 32 + 1.1418$

≈ 1.1766

(c). $H_0: \beta_2 = 0$, $H_1: \beta_2 \neq 0$

(7 points)

$\left| t\text{-statistic} \right| = \left| \frac{0.00042}{6.034 \times 10^{-5}} \right| \approx 6.9481$ (from Table 2)

> 1.96.

$\Longrightarrow$ reject H_0

$\Longrightarrow$ the marginal effect is not constant

(d). $H_0: \beta_2 = \beta_5 = 0$

(9 Points) $H_1: \beta_2 \neq 0$ or $\beta_5 \neq 0$

$$F = \frac{R^2_{ur} - R^2_r}{1 - R^2_{ur}} * \frac{n-k-1}{q}$$

$$\approx \frac{(0.0151 - 0.0017)}{1 - 0.01511} * \frac{233239}{2}$$

$$\approx 1586.7$$

The critical value is $F_{2,233239,1-\alpha} = 3.00$

$\Longrightarrow$ reject H_0

$\Longrightarrow$ The marginal effect is not constant at least for some type of individuals.