The International College of Economics and Finance
Econometrics. Mid-year exam. 2018 October 25. Part 2. (120 minutes).
Suggested Solutions

SECTION A. Answer **ALL** questions **1-3** from this section.

1. [15 marks] You are asked to study the factors that determine employee monthly earnings E_i (in thousands of rubles) of a big corporation (60 observations in total), such as education ED_i, professional experience EXP_i (both in years), age AGE_i (also in years), and AGE_i^2 (in years squared):

$$E_i = \beta_1 + \beta_2 ED_i + \beta_3 EXP_i + \beta_4 AGE_i + \beta_5 AGE_i^2 + u_i; \quad i = 1, 2, \dots, N. \qquad \textbf{(1)}$$

Let estimated equation be

$$\underset{}{\hat{E}_i} = \underset{(800.86)}{-632.01} + \underset{(0.75)}{1.94} ED_i + \underset{(0.86)}{2.17} EXP_i + \underset{(55.95)}{42.36} AGE_i - \underset{(0.97)}{0.74} AGE_i^2 \quad R^2 = 0.18 \qquad \textbf{(1*)}$$

Another estimated equations is

$$\hat{E}_i = \underset{(14.69)}{-27.54} + \underset{(0.69)}{2.10} ED_i + \underset{(0.76)}{2.39} EXP_i \quad R^2 = 0.17 \qquad \textbf{(2*)}$$

(a) [3 marks] Outline briefly how to test whether education matters in determination of the monthly earnings using equation **(1*)**. How would you test whether experience influence positively on the earnings? Is it possible to test whether the earnings of all employees are constant independently of their education level, professional experience and age?

Education: standard t-test for $\hat{\beta}_2$. $H_0 : \beta_2 = 0, H_1 : \beta_2 \neq 0$. $t = \dfrac{1.94}{0.75} = 2.59$. $df = 55$ take $t_{crit}(2tail, 5\%, 50) = 2.009$ - significant at 5%.

Professional experience: one-tailed t-test for $\hat{\beta}_3$. $H_0 : \beta_3 = 0, H_1 : \beta_3 > 0$. Corresponding t-statistic $t = \dfrac{2.17}{0.86} = 2.52$ with the same distribution but now one-tailed critical value $t_{crit}(1tail, 1\%, 50) = 2.403$ should be used.

Earnings are constant: $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0, H_1 : \text{not } H_0 \text{ } or \text{ at least one of } \beta_2, \beta_3, \beta_4, \beta_5 \neq 0.$ Test statistic is $F = \dfrac{R^2/4}{(1-R^2)/(N-5)} = \dfrac{0.18/4}{(1-0.18)/55} = 3.01$, with $F_{crit}(4, 50, 5\%) = 2.56$ - significant at 5%.

(b) [6 marks] Explain the meaning of coefficients $\hat{\beta}_2$ and $\hat{\beta}_3$ in **(1*)**. The coefficient $\hat{\beta}_3$ is a little bigger than $\hat{\beta}_2$. Can you conclude from here that spending a year acquiring the professional experience is more useful for future earnings in the corporation than a year of study? Why or why not? How can you test the hypothesis $\beta_3 = \beta_2$ against $\beta_3 \neq \beta_2$ using F-test? How can you test the hypothesis $\beta_3 = \beta_2$ against $\beta_3 > \beta_2$ using appropriate test?

$\hat{\beta}_2$ **and** $\hat{\beta}_3$: $\hat{\beta}_2 = 1.94$ shows the marginal effect (measured in thousands of rubles) of the year of additional education for the persons of the same age and the same professional experience. $\hat{\beta}_2 = 2.17$ shows the marginal effect (also measured in thousands of rubles) of the year of additional professional experience given age and education unchanged.

As both education and experience are measured in years the coefficients $\hat{\beta}_2$ and $\hat{\beta}_3$ can be compared directly using difference or/and ratio: a year of additional professional experience is more useful for earning than a year of additional study.

To test the hypothesis $\beta_3 = \beta_2$ **against** $\beta_3 \neq \beta_2$ one have to estimate restricted version of equation (1)

$$E_i = \beta_1 + \beta_2(ED_i + EXP_i) + \beta_4 AGE_i + \beta_5 AGE_i^2 + u_i;. \quad \textbf{(3)}$$

and obtain R_R^2 and then compare it with R_U^2 from unrestricted equation **(1*)**, using F-test $F = \dfrac{(R_U^2 - R_R^2)/1}{(1-R_U^2)/55} \sim F(1, 55)$.

To test the hypothesis $\beta_3 = \beta_2$ **against** $\beta_3 > \beta_2$ one should first make reparametrization of the model.

$$E_i = \beta_1 + \beta_2 ED_i + \beta_3 EXP_i + \beta_4 AGE_i + \beta_5 AGE_i^2 + u_i \quad \textbf{(1)}$$

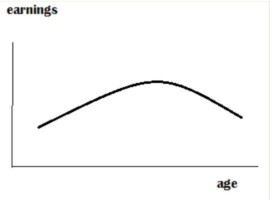
$$E_i = \beta_1 + \beta_2 ED_i + (\beta_3 - \beta_2 + \beta_2) EXP_i + \beta_4 AGE_i + \beta_5 AGE_i^2 + u_i$$

$$E_i = \beta_1 + \beta_2 ED_i + \beta_2 EXP_i + (\beta_3 - \beta_2)EXP_i + \beta_4 AGE_i + \beta_5 AGE_i^2 + u_i$$

$$E_i = \beta_1 + \beta_2 (ED_i + EXP_i) + (\beta_3 - \beta_2)EXP_i + \beta_4 AGE_i + \beta_5 AGE_i^2 + u_i$$

Now we can apply one-tailed t-test to coefficient $(\beta_3 - \beta_2)$ of the variable EXP_i.

(c) [6 marks] Explain the reason for inclusion variable of AGE_i^2 into the model? (*You may use a graphic illustration in addition to the text*). What is marginal effect of age in the model for the employee of 45 years old? Test whether age influence earnings and comment obtained results. Test whether this influence is linear (relationship between earnings and age is a linear function)?



Variable AGE_i^2: inclusion of this variable together with AGE_i makes the relationship on age quadratic. It allows to take into account non-linear effect of age: being too young and too old both is unfavorable for the earnings. There is optimal age when the expected earnings are highest (see picture).

Marginal effect of age. To evaluate marginal effect of age we take the partial derivative of the quadratic relationship of earnings on age at the point $AGE_i = 45$:

$$\frac{\partial}{\partial(AGE_i)} E_i = \frac{\partial}{\partial(AGE_i)} (\hat\beta_1 + \hat\beta_2 ED_i + \hat\beta_3 EXP_i + \hat\beta_4 AGE_i + \hat\beta_5 AGE_i^2) =$$

$= \hat\beta_4 + 2\hat\beta_5 AGE_i = 42.36 - 2 \cdot 0.74 \cdot 45 = -24.24$ thousands of rubles so at the age of 45 years old the marginal effect is negative.

It is not difficult to estimate that the marginal effect of age in this corporation becomes negative approximately at the age of $42.76/2 \cdot 0.74 \approx 29$ years old.

Does age influence earnings? To test this one should test whether coefficients β_4 and β_5 are both equal to zero: $H_0 : \beta_4 = \beta_5 = 0$, H_1 : not H_0 *or* at least one of $\beta_4, \beta_5 \neq 0$. Let us compare determination coefficients from unrestricted equation (1*) and from restricted equation (2*) using F-test $F = \frac{(R_U^2 - R_R^2)/2}{(1 - R_U^2)/55} = \frac{(0.18 - 0.17)/2}{(1 - 0.18)/55} = 0.34$ - obviously insignificant. Possible reason for this insignificance is multicollineatity caused by possible positive correlation between the variable AGE_i and variables ED_i and EXP_i.

Testing whether this influence is linear. It is enough to apply conventional t-test to the coefficient $\hat\beta_5$: $\frac{-0.74}{0.97} = 0.76$. Formally the conclusion is that null hypothesis of linear dependence is not rejected but it contradicts to economic intuition so we should investigate this problem further for extended data to overcome multicollinearity and to get significant result.

2. [15 marks] A researcher is trying to build the best model for the total expenditure on private education $PRIV_t$ in USA (in billions of dollars) for 1993-2017 as a function of disposable personal income DPI_t (also in billions of dollars), relative price index for private education $PRELPRIV_t$ and $TIME_t$ (equal to 1 in 1993, 2 – in 1994 and so on).

(a) [5 marks] First a linear and log-linear models were estimated:

$$\underset{}{\widehat{PRIV}_t} = \underset{(11.81)}{29.72} + \underset{(0.0086)}{0.018} \cdot DPI_t - \underset{(0.14)}{0.38} \cdot PRELPRIV_t + \underset{(0.25)}{0.32} \cdot TIME_t \qquad R^2 = 0.97,\ RSS = 5.54 \qquad \textbf{(1)}$$

$$\widehat{\log(PRIV_t)} = \underset{(1.66)}{6.92} + \underset{(0.0012)}{0.0021} \cdot DPI_t - \underset{(0.020)}{0.071} \cdot PRELPRIV_t + \underset{(0.036)}{0.067} \cdot TIME_t \qquad R^2 = 0.94,\ RSS = 0.11 \qquad \textbf{(2)}$$

Give interpretation to the models and their coefficients (no justification for your interpretation is required). Discuss the significance of coefficient of DPI_t in equation **(2)** using 2-tail and 1-tail tests when needed. What assumptions should be taken to use of 1-tail test? Is it possible to choose between models **(1)** and **(2)** on the base of available information? Why or why not?

Correct interpretation of **(1)** is: increase of disposable personal income by one billion of dollars keeping relative price index and time unchanged leads to the increase in expenditures on private education by 180 millions of dollars.
Increase of relative price index **by one percentage point** keeping disposable personal income and time unchanged leads to the decrease in expenditures on private education by 380 millions of dollars.
Each year the expenditures on private education increases by 320 millions of dollars under assumption that disposable personal income and relative price index are unchanged.
The intercept of equation has no interpretation.
Correct interpretation of **(2)** is: increase of disposable personal income by one billion of dollars keeping relative price index and time unchanged leads to the increase in expenditures on private education by $0.0021 \cdot 100 = 0.21\%$.
Increase of relative price index **by one percentage point** keeping disposable personal income and time unchanged leads to the decrease in expenditures on private education **by** $0.0021 \cdot 100 = 0.21\%$.
Each year expenditures on private education increases by $0.067 \cdot 100 = 6.7\%$ under assumption that other factors included into equation are unchanged.
Significance: t-statistic for coefficient of DPI_t is $0.0021/0.0012 = 1.75$. $t_{crit}(2tail, 5\%, df = 21) = 2.080$ **-** insignificant, but if one assume that DPI_t cannot influence negatively on $PRIV_t$ ($H_0: \beta_{DPI} = 0,\ H_1: \beta_{DPI} > 0$) then we can use 1-tail test with $t_{crit}(1tail, 5\%, df = 21) = 1.721$ **-** significant.
Comparison of equations (1) and (2) is impossible as their dependent variables (and all statistics based on them are different (measured in billions of dollars in **(1)** and in logarithms of billions of dollars in **(2)**) so they have different scale. To compare them we have to apply Zarembka transformation first (see **b**).

(b) [5 marks] The researcher also evaluates variable $PRIVZ_t$ dividing values of $PRIV_t$ by their geometric mean for the whole period $PRIVZ_t = PRIV_t / \sqrt[n]{PRIV_1 \cdot \ldots \cdot PRIV_n}$ and runs new regression

$$\hat{PRIVZ}_t = \underset{(1.18)}{2.97} + \underset{(0.0008)}{0.0018} \cdot DPI_t - \underset{(0.014)}{0.038} \cdot PRELPRIV_t + \underset{(0.025)}{0.032} \cdot TIME_t \qquad \begin{matrix} R^2 = 0.97 \\ RSS = 0.055 \end{matrix} \qquad \textbf{(1*)}$$

Use Box-Cox test to choose between linear and logarithmic specifications of the model. Specify the formula for test statistic, its distribution and corresponding critical values. What is your conclusion? Which model can be selected as the best for further economic and econometric analysis? Explain your choice.

To compare equations **(2)** and **(1*)** we use Box-Cox test based on the comparison of their RSS's: $\chi^2 = (25/2)|\log(0.11/0.055)| = 8.61 > 7.81 = \chi^2_{crit}(5\%, df = 3)$.

Thus linear specification **(1*)** with lower RSS (after Zarembka transformation) is preferred. But it cannot be used for analysis directly as its coefficients are not the same as in original equation (1): if $a = \sqrt[n]{PRIV_1 \cdot \ldots \cdot PRIV_n}$ then

$$PRIVZ_t = \beta_1 + \beta_2 \cdot DPI_t + \beta_3 \cdot PRELPRIV_t + \beta_4 \cdot TIME_t + u_t \qquad (1^*)$$

$$\frac{PRIV_t}{a} = \beta_1 + \beta_2 \cdot DPI_t + \beta_3 \cdot PRELPRIV_t + \beta_4 \cdot TIME_t + u_t \qquad \textbf{and}$$

$$PRIV_t = a\beta_1 + a\beta_2 \cdot DPI_t + a\beta_3 \cdot PRELPRIV_t + a\beta_4 \cdot TIME_t + au_t,$$

and so for further analysis we choose equation **(1)**.

(c) [5 marks] The colleague of the researcher mentioned that equation **(1*)** cannot be compared with log-linear regression **(2)** using Box-Cox test as Zarembka transformation was applied only to linear regression, while equation **(2)** uses untransformed dependent variable, and recommended to apply Zarembka transformation also to **(2)** and use regression **(2*)** instead of **(2)**:

$$\widehat{\log(PRIVZ_t)} = \underset{(1.66)}{4.62} + \underset{(0.0012)}{0.0021} \cdot DPI_t - \underset{(0.020)}{0.071} \cdot PRELPRIV_t + \underset{(0.036)}{0.067} \cdot TIME_t \qquad \begin{matrix} R^2 = 0.94 \\ RSS = 0.11 \end{matrix} \qquad \textbf{(2*)}$$

Comment on her advice.
She also recommended to run double logarithmic regression **(3)**

$$\underset{}{\widehat{\log(PRIV_t)}} = -\underset{(5.19)}{2.38} + \underset{(0.22)}{2.30} \cdot \log(DPI_t) - \underset{(1.13)}{2.24} \cdot \log(PRELPRIV_t) - \underset{(0.02)}{0.02} \cdot TIME_t \quad R^2 = 0.98 \quad RSS = 0.028 \qquad \textbf{(3)}$$

instead of **(2)** and compare it with regression **(1*)** using Box-Cox test. Comment on this advice, do recommended test and make your conclusion. Give interpretation to the coefficients of regression **(3).**

As we can notice all coefficients and their standard errors in **(2*)** are identical to the corresponding coefficients and standard errors of **(2)** except their intercepts. It is no wonder as $\log(\frac{PRIV_t}{a}) = \log(PRIV_t) - \log(a)$ (where $a = \sqrt[n]{PRIV_1 \cdot ... \cdot PRIV_n}$) so equation (1*) differs from equation **(1)** only by the value of intercept. So this advice is not helpful.

The second advice was more interesting. Coefficients of variables DPI_t and $PRELPRIV_t$ in equation (3) are correspondingly income and relative price elasticities of expenditures on private education, and coefficient of $TIME_t$ equal to -0.02 shows that the rate of annual decrease of expenditures on private education is 2% (per year). It is obviously insignificant so its negative sign is obtained probably simply by chance.

Now $\chi^2 = (25/2)\,|\log(0.028/0.055)| = 8.44 > 6.63 = \chi^2_{crit}(1\%, df = 1)$, so now double logarithmic regression having smaller RSS is better.

Regressions (2) and (3) having the same dependent variables can be compared directly by their RSS's (0.028 < 0.11), the regression **(3)** is obviously better.

3. [15 marks] September 2018, elections for the mayor of Moscow were held. The winner was Sergei Sobyanin, his main competitor was Vadim Kumin (KPRF party). A student studying the factors influencing the victory of candidates got exit poll data from one of the polling stations. Each polled voter indicated the candidate for whom he voted (in this polling station all votes were distributed between Sobyanin and Kumin), and also reported how many campaign programs in favor of each of these two candidates (from 0 to 3) he had seen on television. His working hypothesis was that the decision of voters was affected by the number of views of campaign TV programs in favor of each of the candidates. As a result, the student has collected 16 observations: $(S = 0, K = 0, VS_1)$, $(S = 1, K = 0, VS_2)$, …, $(S = 3, K = 3, VS_{16})$, (where S is the number of views for Sobyanin, K is the number of views for Kumin, VS_i is the percentage of votes cast in each group for Sobyanin. For example in the group of voters with 3 TV views in favor of Sobyanin and with 1 TV view in favor of Kumin 96% of voters voted for Sobyanin. He also calculated the total number of TV views for each group $T_i = S_i + K_i$ and calculated sample covariance between S and K: $\mathrm{Cov}(S, K) = 0$.

Then he runs several regressions

$$\hat{VS}_i = 67.35 + 6.85 S_i \qquad R^2 = 0.26 \qquad \textbf{(1)}$$
$$\quad (5.74) \quad (3.06)$$

$$\hat{VS}_i = 79.05 + 6.85 S_i - 7.80 K_i \qquad R^2 = 0.60 \qquad \textbf{(2)}$$
$$\quad (5.60) \quad (2.34) \quad (2.34)$$

$$\hat{VS}_i = 79.05 + 14.65 S_i - 7.80 T_i \qquad R^2 = 0.60 \qquad \textbf{(3)}$$
$$\quad (5.6) \quad (3.30) \quad (2.34)$$

(a) [5 marks] Give interpretation to the coefficients of the variable S_i in equations **(1-3)**. Test the significance of the coefficients of these regressions and the significance of the equations as a whole. (*Do not compare coefficients or other characteristics of equations 1-3, you will be asked to do this later in **b** and **c***)

Interpretation.

In regression (1) the coefficient of S_i shows how many percentage points to the vote result for Sobyanin adds one additional view of the TV program in favor of Sobyanin.

In regression (2) the interpretation is almost the same but this effect is evaluated under assumption that the number of views of the TV program in favor of Kumin remains constant (so it is pure marginal effect of 'in favor of Sobyanin views').

In regression (3) the interpretation is very much the same as we assume the total number of TV views T_i constant.

To test the significance of the coefficients

	Intercept	Slope coefficient
Regression 1	$t = \frac{67.35}{5.74} = 11.73$	$t = \frac{6.85}{3.06} = 2.24$
Critical value	$t_{crit}(2-tail, df = 14, 1\%) = 2.97$	$t_{crit}(2-tail, df = 14, 5\%) = 2.145$
Conclusion	Significant at 1% level	Significant at 5% level

	Intercept	First slope coefficients	Second slope coefficient
Regression 2	$t = \frac{79.05}{5.60} = 14.12$	$t = \frac{6.85}{2.34} = 2.93$	$t = \frac{-7.80}{2.34} = -3.33$
Critical value	$t_{crit}(2tail, 13, 1\%) = 3.012$	$t_{crit}(2tail, 13, 5\%) = 2.160$	$t_{crit}(2tail, 13, 1\%) = 3.01$
Conclusion	Significant at 1% level	Significant at 5% level	**Significant at 1% level**

	Intercept	First slope coefficients	Second slope coefficient
Regression 3	- - -	$t = \frac{14.65}{3.30} = 4.44$	- - -
Critical value	- - -	$t_{crit}(2tail, 13, 1\%) = 3.012$	- - -
Conclusion	- - -	**Significant at 1% level**	- - -

Test the significance of the equation as a whole. Equation **(1)** does not require F-test as for simple regression F test and t-test for the slope are equivalent.

For equations **(2)** and **(3)** $F = \frac{R^2/2}{(1-R^2)/13} = \frac{0.6/2}{(1-0.6)/13} = 9.75$ while $F_{crit}(2, 13, 1\%) = 6.7$ so both equations are significant at 1%.

(b) [6 marks] Why the intercepts in equations **(2)** and **(3)** are identical while in **(1)** it is different? Explain why the coefficient of the variable T_i in equations **(3)** is identical to the coefficient of the variable K_i in equation **(2)**? Explain why the coefficient of the variable S_i in equations **(3)** is bigger than in equation **(2)**? Explain why the coefficients of the variable S_i in equations **(1)** and **(2)** are identical? (*You may use the formulas for estimators of regression coefficients or any other method for your explanations*).

Write down equations in theoretical form

$$VS_i = \beta_1^0 + \beta_2^0 S_i + u_i \qquad (1)$$

$$VS_i = \beta_1 + \beta_2 S_i + \beta_3 K_i + v_i \quad (2)$$

$$VS_i = \beta_1^* + \beta_2^* S_i + \beta_3^* T_i + w_i \quad (3)$$

The intercepts in (2) and (3) shows value of VS_i for the group of voters with zero TV views (both in favor of Sobyanin and Kumin). In regression (1) Kumin views are not under control while Sobyanin views are zero, so no wonder that the percentage of votes for Sobyanin here has become lower.

β_3 in (2) shows the marginal effect of K_i given S_i is fixed, but if K_i in (3) is fixed then marginal effect of T_i is fully achieved at the expense of K_i so the coefficients β_3 and β_3^* are really identical.

Now consider coefficient β_2 in (2). It shows marginal effect of S_i given K_i is fixed. β_2^* in (3) also shows marginal effect of S_i but now given T_i is fixed, as $T_i = S_i + K_i$ this means that when S_i increases by one view K_i should decrease by one view. So β_3^* should be equal to the difference of coefficients $\beta_2 - \beta_3$. And we see that $14.65 = 6.85 - (-7.80)$.

As for the equality of coefficients in simple and multiple regressions several explanations are possible. The simplest one is following. As it was mentioned above β_2 in (2) shows marginal effect of S_i given K_i is fixed. In simple regression model (1) β_2^0 also takes into account the interaction between S_i and K_i. But from covariance matrix we can see that this interaction (covariance and so correlation) is zero. So coefficient of S_i does not change.

The same also can be derived formally.

$$\hat{\beta}_2^0 = \frac{\mathrm{Cov}(VS,S)}{\mathrm{Var}(S)}, \quad \text{and} \quad \hat{\beta}_2 = \frac{\mathrm{Cov}(VS,S)\,\mathrm{Var}(K) - \mathrm{Cov}(VS,K)\,\mathrm{Cov}(S,K)}{\mathrm{Var}(S)\,\mathrm{Var}(K) - (\mathrm{Cov}(S,K))^2} =$$

$$= \frac{\mathrm{Cov}(VS,S)\,\mathrm{Var}(K) - \mathrm{Cov}(VS,K)\cdot 0}{\mathrm{Var}(S)\,\mathrm{Var}(K) - (0)^2} = \frac{\mathrm{Cov}(VS,S)\,\mathrm{Var}(K)}{\mathrm{Var}(S)\,\mathrm{Var}(K)} = \frac{\mathrm{Cov}(VS,S)}{\mathrm{Var}(S)} = \hat{\beta}_2^0$$

In the mathematical language this can be also expressed in the following way: from the construction of the sample we can see that variables S_i and K_i are **orthogonal** (their scalar product is zero) so they are not correlated by definition independently of the values of VS_i. Multiple regression with such variables is usually called in econometrics as orthogonal regression – its coefficients are always the same as in partial simple regression models.

(c) [4 marks] Explain why determination coefficients R^2 in equations **(2)** and **(3)** are identical, while in equation **(1)** R^2 is about three times less. Explain why in equation **(2)** the standard errors of both slope coefficients are identical. Suggest probable reason why the standard error of the coefficient for the variable S_i in equation **(2)** is less than corresponding standard error in equation **(1)**.

R^2 always not decreases with the inclusion new variables into equation so no wonder that R^2 in (1) is less than in (2) and (3).

Both equations (2) and (3) are based on the same data. Substituting $T_i = S_i + K_i$ into (3)

$$VS_i = \beta_1^* + \beta_2^* S_i + \beta_3^* T_i + w_i \quad \textbf{(3)}$$

$$VS_i = \beta_1^* + \beta_2^* S_i + \beta_3^* (S_i + K_i) + w_i = \beta_1^* + (\beta_2^* + \beta_3^*) S_i + \beta_3^* K_i + w_i$$

We can see this equation matches exactly **(2)**

$$VS_i = \beta_1 + \beta_2 S_i + \beta_3 K_i + v_i \quad \textbf{(2)}$$

Standard errors of coefficients in equation (2) are evaluated by formulas

$s.e.(\hat{\beta}_2) = \sqrt{\dfrac{s_u^2}{n\,\mathrm{Var}(S)}} \cdot \sqrt{\dfrac{1}{1 - r_{S,K}^2}}$ and $s.e.(\hat{\beta}_3) = \sqrt{\dfrac{s_u^2}{n\,\mathrm{Var}(K)}} \cdot \sqrt{\dfrac{1}{1 - r_{S,K}^2}}$. From the construction of the sample both variables S_i and K_i takes the same values 1, 2, 3, and 4 with equal frequency ¼ so their sample variances are equal $\mathrm{Var}(S) = \mathrm{Var}(K)$.

Probable reason why the standard error of the coefficient for the variable S_i in equation (2) is less than corresponding standard error in equation (1) can be following.

Standard error of coefficient of S_i in equation (2) is evaluated by formula

$s.e.(\hat{\beta}_2) = \sqrt{\frac{s_u^2}{n\,\mathrm{Var}(S)}} \cdot \sqrt{\frac{1}{1 - r_{S,K}^2}}$ but $r_{S,K}^2 = 0$, so $s.e.(\hat{\beta}_2) = \sqrt{\frac{s_u^2}{n\,\mathrm{Var}(S)}} = \sqrt{\frac{RSS(regr.2)}{n\,\mathrm{Var}(S)}}$. It has the same form as the standard error for simple linear regression $s.e.(\hat{\beta}_2^0) = \sqrt{\frac{s_u^2}{n\,\mathrm{Var}(S)}} = \sqrt{\frac{RSS(regr.1)}{n\,\mathrm{Var}(S)}}$. Regression (1) is an restricted version of regression (2) so $RSS(regr.1) \geq RSS(regr.2)$ so $s.e.(\hat{\beta}_2^0) \geq s.e.(\hat{\beta}_2)$.

Bonus question: what drawbacks can you indicate in the student's study?

1) The sample is too small as the student use aggregated data to get percentages.
2) The sample is not representative, as it is not random. Only data from one polling station was used where all voters cast their votes only to two candidates. So the results can be biased.
3) In econometrics (in OLS) all observations are treated as having equal weight. So the observation (S=3, K=0) has the same 'weight' as (S=0, K=3) while in reality the first one probably has much greater frequency.

SECTION B. Answer **ONE** question from this section (**4 OR 5**).

4. [30 marks] At a lecture on econometrics the professor said to the students that OLS estimator for the coefficient α in a simple regression model without constant $Y_t = \alpha X_t + u_t$, $t = 1, 2, \ldots, T$ is given by $\hat{\alpha}_{OLS} = \frac{\sum_{t=1}^{T} X_t Y_t}{\sum_{t=1}^{T} X_t^2}$. But this estimator is not the only one, there is another estimator of α, $\hat{\alpha}_1 = \frac{\overline{Y}}{\overline{X}} = \frac{\frac{1}{T}\sum Y_t}{\frac{1}{T}\sum X_t}$.

(a) [8 marks] Show that $\hat{\alpha}_1 = \frac{\overline{Y}}{\overline{X}}$ is also an unbiased estimator of α in $Y_t = \alpha X_t + u_t$. What assumptions are needed for this result being valid? Is $\overline{Y} / \overline{X}$ also a good estimator for β_2 in $Y_t = \beta_1 + \beta_2 X_t + u_t$? Explain.

$(\bar{Y}/\bar{X})$. $\hat{\alpha} = \frac{\bar{Y}}{\bar{X}} = \frac{\frac{1}{T}\sum Y_t}{\frac{1}{T}\sum X_t} = \frac{\sum(\alpha X_t + u_t)}{\sum X_t} = \alpha + \frac{\sum u_t}{\sum X_t}$, hence $E(\hat{\alpha}) = \alpha + \frac{\sum E(u_t)}{\sum X_t} = \alpha$ under assumptions $E(u_t) = 0$ and X_t is non-stochastic.

If the regression equation now contains a constant term then the method is no longer appropriate unless the constant term $=0$. If $\beta_1 \neq 0$ then the estimator $\frac{\bar{Y}}{\bar{X}}$ is biased:

$\hat{\alpha} = \frac{\bar{Y}}{\bar{X}} = \frac{\frac{1}{T}\sum Y_t}{\frac{1}{T}\sum X_t} = \frac{\sum(\beta_1 + \beta_2 X_t + u_t)}{\sum X_t} = \beta_2 + \frac{\beta_1}{\sum X_t} + \frac{\sum u_t}{\sum X_t}$ so

$E(\hat{\alpha}) = \beta_2 + \frac{\beta_1}{\sum X_t} + \frac{\sum E(u_t)}{\sum X_t} = \beta_2 + \frac{\beta_1}{\sum X_t}$ under the same assumptions.

The same can be shown also graphically: the estimated regression line goes through the origin and the point ($\bar{X}, \bar{Y}$) and therefore always has the slope $\hat{\alpha} = \frac{\bar{Y}}{\bar{X}}$, so if true regression $Y_t = \beta_1 + \beta_2 X_t + u_t$ has different slope β_2 the estimator will be biased.

(b) [7 marks] Derive population variance of the estimator $\hat{\alpha}_1 = \frac{\bar{Y}}{\bar{X}}$. What assumptions are used to derive your result?

Let us derive variance of $\hat{\alpha}_1$.

$$\text{var}(\hat{\alpha}_1) = \text{var}\left(\frac{\frac{1}{T}\sum_{t=1}^{T} Y_t}{\frac{1}{T}\sum_{t=1}^{T} X_t}\right) = \frac{\sum_{t=1}^{T}\text{var}\, Y_t}{(\sum_{t=1}^{T} X_t)^2} = \frac{\sum_{t=1}^{T}\sigma^2}{(\sum_{t=1}^{T} X_t)^2} = \frac{T\sigma^2}{(\sum_{t=1}^{T} X_t)^2} = \frac{1}{T}\frac{T^2\sigma^2}{(\sum_{t=1}^{T} X_t)^2} = \frac{1}{T}\frac{\sigma^2}{\left(\frac{\sum_{t=1}^{T} X_t}{T}\right)^2} = \frac{1}{T}\frac{\sigma^2}{(\bar{X})^2} .$$

Important comments: $\text{var}(\sum_{t=1}^{T} Y_t) = \sum_{t=1}^{T}\text{var}\, Y_t + 2\sum_{t \neq s}\text{cov}(Y_t, Y_s) = \sum_{t=1}^{T}\text{var}\, Y_t = T\sigma_u^2$ under assumption $\text{var}\, Y_t = \sigma_u^2$.

For $s \neq t$ $\text{cov}(Y_t, Y_s) = \text{cov}(\alpha X_t + u_t, \alpha X_s + u_s) =$

$= \text{cov}(\alpha X_t, \alpha X_s) + \text{cov}(\alpha X_t, u_s) + \text{cov}(u_t, \alpha X_s) + \text{cov}(u_t, u_s) = 0 + 0 + 0 + \text{cov}(u_t, u_s) = \text{cov}(u_t, u_s) = 0$

under assumptions that X_t is non-stochastic and $\text{cov}(u_t, u_s) = 0$, $s \neq t$.

(c) [8 marks] Giving a lecture an absent-minded professor by mistake wrote on a board the formula $\hat{\alpha} = \overline{Y_t / X_t}$ instead of $\hat{\alpha} = \bar{Y}/\bar{X}$ ($\overline{Y_t / X_t} = \frac{1}{T}\sum_{t=1}^{T}(Y_t / X_t)$). Does it change the conclusion that the estimator is unbiased? Show that population variance of $\hat{\alpha}_2 = \overline{Y_t / X_t}$ is given by the expression $\text{var}(\hat{\alpha}_2) = \frac{\sigma^2}{T^2}\sum_{t=1}^{T}(1/X_t^2)$. State the assumptions needed for your work.

$E(\hat{\alpha}_2) = E(\frac{1}{T}\sum_{t=1}^{T}\frac{Y_t}{X_t}) = \frac{1}{T}E(\sum_{t=1}^{T}\frac{Y_t}{X_t}) = \frac{1}{T}\sum_{t=1}^{T}\frac{E(Y_t)}{X_t} = \frac{1}{T}\sum_{t=1}^{T}\frac{\alpha X_t}{X_t} = \frac{1}{T}\sum_{t=1}^{T}\alpha = \frac{T\alpha}{T} = \alpha$.

Hence $\hat{\alpha}_2 = \overline{Y_t / X_t}$ is an unbiased estimator as long as $E(u_t) = 0$ and X_t is non-stochastic.

$$\mathrm{var}(\hat{\alpha}_2) = \mathrm{var}(\frac{1}{T}\sum_{t=1}^{T}(Y_t / X_t)) = \frac{1}{T^2}\mathrm{var}(\sum_{t=1}^{T}(Y_t / X_t)) = \frac{1}{T^2}\sum_{t=1}^{T}\frac{\mathrm{var}(Y_t)}{X_t^2} = \frac{1}{T^2}\sum_{t=1}^{T}\frac{\sigma^2}{X_t^2} = \frac{\sigma^2}{T^2}\sum_{t=1}^{T}\frac{1}{X_t^2}.$$

Important comments: $\mathrm{var}(\sum_{t=1}^{T}\frac{Y_t}{X_t}) = \sum_{t=1}^{T}\frac{\mathrm{var}\, Y_t}{X_t^2} + 2\sum_{t \neq s}\mathrm{cov}(\frac{Y_t}{X_t}, \frac{Y_s}{X_s}) = \sum_{t=1}^{T}\frac{\mathrm{var}\, Y_t}{X_t^2} = \sigma_u^2\sum_{t=1}^{T}\frac{1}{X_t^2}$.

For $s \neq t$ $\mathrm{cov}(\frac{Y_t}{X_t}, \frac{Y_s}{X_s}) = \mathrm{cov}(\frac{\alpha X_t + u_t}{X_t}, \frac{\alpha X_s + u_s}{X_s}) =$

$$= \frac{1}{X_t X_s}[\mathrm{cov}(\alpha X_t, \alpha X_s) + \mathrm{cov}(\alpha X_t, u_s) + \mathrm{cov}(u_t, \alpha X_s) + \mathrm{cov}(u_t, u_s)] = \frac{0+0+0+\mathrm{cov}(u_t, u_s)}{X_t X_s} = \frac{\mathrm{cov}(u_t, u_s}{X_t X_s}$$

These transformations used the assumptions that $\sum_{t=1}^{T} u_t = 0$, X_t is non-stochastic and $\mathrm{var}\, Y_t = \sigma_u^2$.

(d) [7 marks] Which of the two estimators $\hat{\alpha}_1 = \overline{Y} / \overline{X}$ and $\hat{\alpha}_2 = \overline{Y_t / X_t}$ is more efficient? Explain what you understand by efficiency.

Hint: use inequality for the four means (quadratic, arithmetic, geometric, harmonic:

$$\boxed{\sqrt{\frac{1}{T}\sum_{t=1}^{T}x_t^2} \geq \frac{\sum_{t=1}^{T}x_t}{T} \geq \sqrt[T]{x_1 \cdot x_2 \cdot \ldots \cdot x_T} \geq \frac{T}{(\sum_{t=1}^{T}\frac{1}{x_t})}}$$

More efficient' for two unbiased estimators means that the estimator has the smaller variance. Let us compare two variances (assuming all X_t positive)

$\mathrm{var}(\hat{\alpha}_1) = \mathrm{var}(\frac{\overline{Y}}{\overline{X}}) = \frac{1}{T}\frac{\sigma^2}{(\overline{X})^2}$ while $\mathrm{var}(\hat{\alpha}_2) = \mathrm{var}(\frac{1}{T}\sum_{t=1}^{T}(Y_t / X_t)) = \frac{\sigma^2}{T^2}\sum_{t=1}^{T}\frac{1}{X_t^2}$.

We will show that the first variance is less (in general).

The inequality $\mathrm{var}(\hat{\alpha}_1) \leq \mathrm{var}(\hat{\alpha}_2)$ that is $\frac{1}{T}\frac{\sigma^2}{(\overline{X})^2} \leq \frac{\sigma^2}{T^2}\sum_{t=1}^{T}\frac{1}{X_t^2}$ is equivalent to $\frac{1}{(\overline{X})^2} \leq \frac{1}{T}\sum_{t=1}^{T}\frac{1}{X_t^2}$.

Let $\frac{1}{X_t} = x_t$ so $X_t = \frac{1}{x_t}$ then $\frac{1}{(\overline{X})^2} = \frac{1}{(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{x_t})^2} = \frac{T^2}{(\sum_{t=1}^{T}\frac{1}{x_t})^2}$. **(*)**

Note that inequality for means (harmonic mean $\leq$quadratic mean) being squared takes the form $\frac{T^2}{(\sum_{t=1}^{T}\frac{1}{x_t})^2} \leq \frac{1}{T}\sum_{t=1}^{T}x_t^2$. Applying it to **(*)** we get $\frac{1}{(\overline{X})^2} = \frac{1}{(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{x_t})^2} = \frac{T^2}{(\sum_{t=1}^{T}\frac{1}{x_t})^2} \leq \frac{1}{T}\sum_{t=1}^{T}x_t^2 = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{X_t^2}$. QED.

Question 5. [30 marks] A researcher has data on income *Y*, capital *K* and labor *L* indices for 1991-2017 related to the economy of some developing country (time series specific problems are out of consideration in this question). She estimated relationship between these variables using different production functions.

(a) [8 marks] First she runs a linear model

$$\hat{Y}_t = \underset{(14.0)}{10.83} + \underset{(0.025)}{0.22} \cdot K_t + \underset{(0.12)}{0.72} \cdot L_t \qquad R^2 = 0.95 \qquad RSS = 3096.4 \qquad \textbf{(1)}$$

Noticing that the sum of the slope coefficients is close to unity she decided to test the restriction $\beta_K + \beta_L = 1$, and decided to run the restricted regression

$$Y = \alpha + \theta \cdot K + (1 - \theta) \cdot L + u, \qquad \textbf{(2)}$$

How equation **(2)** could be estimated using OLS (you may use results of this estimation $R^2 = 0.85$ and $RSS = 3137.9$)? How to test the restriction? Is it possible to run F-test, based on comparison of determination coefficients of restricted and unrestricted regressions? Why or why not? What is economic meaning of this restriction (if any)? Would you recommend to accept the restriction?

To estimate equation (2) it is sufficient to rearrange terms:
First

$$Y = \alpha + \theta \cdot K + L - \theta \cdot L + u$$

and then

$$Y - L = \alpha + \theta \cdot (K - L) + u \qquad \mathbf{(2^*)}$$

which could be estimated using OLS to get R_R^2 and RSS_R. Only RSS's can be used here as **(1)** and **(2*)** have different dependent variables so they have different TSS's and so their determination coefficients are incomparable.

F-test based on comparison of RSS's is $F = \dfrac{RSS_R - RSS_U}{RSS_U} = \dfrac{(3137.9 - 3096.4)/1}{3096.4/24} = 0.32$ - obviously insignificant.

Equation **(1)** is a linear production function, the coefficients show partial marginal effect of capital and labor on income. But the restriction itself is nonsense from economic point of view and has nothing to do with the condition of constant returns to scale. The sum of coefficients here is close to unity simply by chance. Nevertheless imposing any restriction when it is true helps to make estimates more efficient.

(b) [7 marks] Then the researcher runs auxiliary regression **(1a)**,

$$\hat{Y}_t = -51.36 + 0.48 \cdot K_t + 1.30 \cdot L_t - 0.003 \cdot (\hat{Y}_t^*)^2 \qquad R^2 = 0.96 \qquad \mathbf{(1a)}$$
$$(26.6)\ \ (0.10) \qquad (0.24) \qquad (0.001) \qquad RSS = 2375$$

where $(\hat{Y}_t^*)^2$ are squared estimated values of Y_t from the equation **(1)**: $\hat{Y}_t^* = 10.83 + 0.22 \cdot K_t + 0.72 \cdot L_t$
Comment on the meaning of researcher's actions, run the appropriate test and draw the conclusion. What implicit economic assumptions are used in the specification of equation **(1)**? To what extent can they be considered relevant to economic theory and practice? How to improve the specification of the equation **(1)**?

Equation (1a) is an auxiliary equation used for testing whether the specification of the equation (1) is correct (Ramsey test). Following the logic of Ramsey test (see the lectures) $t = \dfrac{-0.003}{0.001} = -3$, so $|t| = 3$ while $t_{crit}(2\text{tail}, 24, 1\%) = 2.797$ - reject $H_0 : \beta_{\hat{Y}_t^*} = 0$.

As the coefficient by the variable $(\hat{Y}_t^*)^2$ is significant one could suspect that some kind of nonlinearity may be present in the data.
The equation **(1)** uses linear specification for the production function. Using linear function is based on the implicit assumption that marginal effects of capital and labor on income are constant what is hardly be true. Moreover these effects are additive in equation what implies that these two factors are infinitely interchangeable, which obviously is not true.
In econometrics linear production functions are commonly used for the data describing growth rates like this

$$y = \alpha + \theta \cdot k + \gamma \cdot l + u$$

where y - growth rate of income Y, k - growth rate of capital K, l - growth rate of labor L
Better use multiplication of capital and labor instead of their addition, and use power coefficients. The production function in this form is well known Cobb-Douglas production function. In double logarithmic specification we can see it in **(c)**.

(c) [8 marks] Next she runs two logarithmic regressions

$$\log(Y) = 0.48 + 0.37 \cdot \log(K) + 0.53 \cdot \log(L) \qquad R^2 = 0.97 \qquad \mathbf{(3)}$$
$$(0.39)\ (0.05) \qquad (0.12) \qquad RSS = 0.0763$$

and

$$\log(\hat{Y/L}) = -0.0050 + 0.33 \cdot \log(K/L) \qquad R^2 = 0.82 \qquad \mathbf{(4)},$$
$$(0.0183 \quad (0.03) \qquad RSS = 0.0811$$

Give the interpretation of both models and their parameters. Prove that equation **(4)** is a restricted version of the equation **(3)**. What is the restriction? Is it significant? What equation would you choose and why?

c) Equation **(3)** is a standard specification of Cobb-Douglas function. The coefficients of this equation are capital elasticity of income and labor elasticity of income. It means that 1% increase of *K* implies 0.37% increase of income *Y*, while 1% increase of *L* implies 0.53% increase of income *Y*.
Both coefficients are significant at 1% significance level.
The variables of this equation are different from the variables of equation (3). Equation (4) describes the dependence of labor productivity from the capital-labor ratio. From the technical point of view equation (4) simply allows to test the restriction $\alpha+\beta=1$ for equation $Y=AK^{\alpha}L^{\beta}v$ (constant returns to scale condition).
It means that if *K* and *L* increase *n* times, then *Y* also will increase n times (property of homogeneous function).
Let us show that equation (4) is a restricted version of (3). Equation $\log(Y)=\alpha+\beta_K\cdot\log(K)+\beta_L\cdot L+u$ corresponds to the function $Y=y^{\alpha}\cdot K^{\beta_K}\cdot L^{\beta_L}e^{u}$ If $\beta_K+\beta_L=1$ then $Y=y^{\alpha}\cdot K^{\beta_K}\cdot L^{1-\beta_K}e^{u}$, Dividing both parts by L we get $Y/L=y^{\alpha}\cdot(K/L)^{\beta_K}e^{u}$ and taking logarithms we get $\log(Y/L)=\alpha+\beta_K\cdot\log(K/L)+u$ which corresponds to (4).
To test this restriction one should evaluate *F*-statistics:

$$F=\frac{(RSS_{restricted}-RSS_{unrestricted})/number\,of\;restr.}{RSS_{unrestricted}/(d.f.of\;unrestricted)}=\frac{(0.0811-0.0763)/1}{0.0763/(27-3)}=1.51$$

what is less than $F(crit.,5\%,1,24)=4.26$, so the restriction cannot be rejected.
So we should choose an equation (3) to get more efficient estimation.
The values of elasticity are close to those what are typical for developing countries (quite a big value of labor elasticity of income).

(d) [7 marks] Now the researcher introduces time (*T* equal to 1 in the first year of the period, 2 – in the second and so on), and estimates Cobb-Douglas (CD) production function using two different methods:
OLS applied to linearized model:

$$\widehat{\log(Y)}=2.53-\underset{(1.17)}{}\;\underset{(0.24)}{0.067}\cdot\log(K)+\underset{(0.11)}{0.51}\cdot\log(L)+\underset{(0.016)}{0.03}\cdot T\qquad R^2=0.97,\; RSS=0.0664\qquad \textbf{(5-OLS)}$$

NLS (Non-Linear Least Squares) applied to original model (all estimates except capital elasticity are significant) :

$$\hat{Y}=e^{3.8}\cdot K^{-0.046}\cdot L^{0.086}\cdot e^{0.008T}\qquad R^2=0.97\qquad \textbf{(5-NLS)}$$

Comment on parameters of both equations, paying special attention to the coefficients of variables K and *T*. Why estimates of equations **(5-OLS)** and **(5-NLS)** are different?

This equation without logarithms is $\hat{Y}=e^{2.53}K^{0.067}L^{0.51}e^{0.03T}$. The coefficient of *T* multiplied by 100 is income growth rate induced by the other factors than capital and labor (they are supposed to be fixed).
Usually this growth rate (here 3%) is attributed to technological progress. Both equations are Cobb-Douglas (CD) production functions that take into account technological progress.
The coefficient of *K* has wrong sign and insignificant. This can be explained in various ways: maybe capital is not limiting factor in the economy under consideration, but most probable explanation is that *K* and *T* are strongly correlated which leads to multicollinearity.
The estimates of both equations are different as two different methods are applied to them. Both methods aim to find the least sum of squares, but all these squares are taken using different functions with totally different disturbance terms: multiplicative disturbance term in (5-OLS): $\log Y=\beta_1+\beta_2\log K+\beta_3\log L+\beta_4+u$ which equivalent to $Y=e^{\beta_1}K^{\beta_2}L^{\beta_3}e^{\beta_4 T}e^{u}$ with multiplicative disturbance term $v=e^{u}$, and additive disturbance term in (5-NLS): $Y=e^{\beta_1}K^{\beta_2}L^{\beta_3}e^{\beta_4 T}+\varpi$. As we can see, the difference in the form of inclusion of a random term causes significant changes in the estimates obtained.