The International College of Economics and Finance
Econometrics – 2021-2022. Midterm exam. 2021 October 21.
Part 2. Free Response Questions (1 hour 30 minutes)

SECTION A

Answer **ALL** questions from this section (questions **1-2**).

Question 1. (25 marks)

Let a regression equation be:

$$\log Y_t = \beta_1 + \beta_2 \log K_t + \beta_3 \log L_t + u_t;\ t = 1,2,...,T \qquad \textbf{(1)}$$

where Y_t is income, K_t is capital and L_t is labor (all are index numbers).

(a) [12 marks] □ The researcher believes that for the data under consideration K_t influence Y_t; how this belief can be tested?

$H_0 : \beta_2 = 0$ against $H_a : \beta_2 \neq 0$. Standard two-tailed t test $t = \dfrac{\hat{\beta}_2}{\text{s.e.}(\hat{\beta}_2)}$. If $|t| > t_{crit.}^{\alpha}(T-3)$ we reject H_0 at the specified significance level α

□ The researcher believes that for the data under consideration the influence of K_t on Y_t is strictly positive; how he can test his belief? Give a pair of hypotheses for the formal test and describe how to do it.

$H_0 : \beta_2 = 0$ against $H_a : \beta_2 > 0$. Then the same t-statistics $t = \dfrac{\hat{\beta}_2}{\text{s.e.}(\hat{\beta}_2)}$ should be compared with one tailed critical value, if $t > t_{crit.}^{\alpha}(T-3, one\ tail)$ the null hypothesis is rejected

□ How to test that jointly β_2 and β_3 are significantly different from zero.

To test $H_0 : \beta_2 = \beta_3 = 0$ against H_A: at least one of β_2, β_3 is not zero, we use F test: $F = \dfrac{R^2/2}{(1-R^2)/(T-3)} \sim F(2, N-3)$. If $F > F_{crit}^{\alpha}(2, T-3)$ the null is rejected.

□ Assume both β_2 and β_3 are significant. Does it follow from here that β_2 and β_3 are jointly significant?
No, the significance of any test is a result of calculations on the base of random sample, so any combinations of results of conventional t-tests for significance of coefficients and F-test for the their joint significance are possible.

(b) [13 marks] □ How to test the restricion $\beta_2 + \beta_3 = 1$?

To test $H_0 : \beta_2 + \beta_3 = 1$ or $H_0 : \beta_3 = 1 - \beta_2$ against $H_A : \beta_2 + \beta_3 \neq 1$ we use F-test, that allows to compare unrestricted (*U*) equation

$$\log Y_t = \beta_1 + \beta_2 \log K_t + \beta_3 \log L_t + u_t \qquad \text{(U)}$$

with the restricted (*R*) one

$$\log Y_t = \beta_1 + \beta_2 \log K_t + (1-\beta_2) \log L_t + u_t \qquad \text{(R)}$$

So running regression (U) (OLS) we get RSS_U

$$\log \hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 \log K_t + \hat{\beta}_3 \log L_t \quad \rightarrow \quad RSS_U$$

The estimation of restricted equation directly by OLS is impossible

$$\log Y_t = \beta_1 + \beta_2 \log K_t + (1-\beta_2) \log L_t + u_t$$

for estimation one should do reparametrisation of this equation in the following way

$$\log Y_t - \log L_t = \beta_1 + \beta_2 (\log K_t - \log L_t) + u_t$$

So running last regression (OLS) we get RSS_R

Now calculate $F = \frac{(RSS_R - RSS_U)/1}{RSS_U/(T-3)} \sim F(1, T-3)$. If $F > F_{crit}(1, T-3, \alpha\%)$ then H_0 is rejected.

□ Explain whether a test based on R^2 can be used here.

Here we cannot use $F = \frac{(R_U^2 - R_R^2)/1}{(1-R_U^2)/(T-3)}$, because it involves $R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$. RSS is the same for the model with dependent variable $\log Y_t$ and $\log Y_t - \log L_t$, but their TSS 's are obviously different.

□ Show mathematically that function $Y_i(K_i, L_i)$ defined by **(1)** with the restriction $\beta_2 + \beta_3 = 1$ exhibit conctant returns to scale.

Definition: the property $f(ax_1, ax_2) = af(x_1, x_2)$ is called **constant returns to scale**. From $\log Y_t = \beta_1 + \beta_2 \log K_t + \beta_3 \log L_t + u_t;\ t = 1,2,...,T$ we get $Y_t = e^{\beta_1} K_t^{\beta_2} L_t^{\beta_3} e^{u_t}$. So

$$Y_t(aK_t, aL_t) = e^{\beta_1}(aK_t)^{\beta_2}(aL_t)^{\beta_3} e^{u_t} = (a^{\beta_2} a^{\beta_3}) e^{\beta_1}(K_t)^{\beta_2}(L_t)^{\beta_3} e^{u_t} =$$
$$= (a^{\beta_2+\beta_3}) \cdot e^{\beta_1}(K_t)^{\beta_2}(L_t)^{\beta_3} e^{u_t} = ae^{\beta_1}(K_t)^{\beta_2}(L_t)^{\beta_3} e^{u_t} = aY_t(K_t, L_t)$$

□ How to test the restriction $\beta_2 + \beta_3 = 1$ against $\beta_2 + \beta_3 < 1$?

F-test here is not applicable as F-distribution is based on the sum of squares that does not disriminate deviations to the left and to the right. To test $H_0: \beta_2 + \beta_3 - 1 = 0$ against $H_A: \beta_2 + \beta_3 > 1$ or $H_A: \beta_2 + \beta_3 < 1$ we have to use t-distribution after reparametrization of restricted equation.

$$Y_t = \beta_1 + (\beta_2 + \beta_3 - 1)X_{2t} - \beta_3 X_{2t} + X_{2t} + \beta_3 X_{3t} + u_t;\ t = 1,2,...,T$$

To estimate this equation we need to rearrange terms

$$Y_t - X_{2t} = \beta_1 + (\beta_2 + \beta_3 - 1)X_{2t} + \beta_3(X_{3t} - X_{2t}) + u_t;\ t = 1,2,...,T$$

Now test the significance of the coefficient of the variable X_{2t} using **one-tailed test**.

If $t_{\beta_2 + \hat{\beta}_3 - 1} < -t(crit.one\ tail)$ then we reject $H_0: \beta_2 + \beta_3 - 1 = 0$ in favor of $H_A: \beta_2 + \beta_3 < 1$

Question 2. (25 marks)

A relationship between aggregate expenditure on kitchen appliances in the USA (in billions of dollars) and aggregate personal disposable income (also in billions of dollars) in 2000-2019 is investigated. For this purpose, several regressions are fitted.

a) [12 marks] The regression of the expenditure on kitchen appliances KIT on the personal disposable income, gives the following result:

$$\hat{KIT} = -1.77 + 0.0054 \cdot DPI \qquad RSS = 35,95 \qquad R^2 = 0.96 \qquad (1)$$
$$\log(\hat{KIT}) = 1.44 + 0.00038 \cdot DPI \qquad RSS = 0.37 \qquad R^2 = 0.93 \qquad \mathbf{(2)}$$

The Zarembka transformation is now performed: a new variable $KITZ$ is generated instead of KIT and the following equation is obtained:

$$\hat{KITZ} = -0.12 + 0.00038 DPI \qquad RSS = 0.17 \qquad R^2 = 0.96 \qquad \mathbf{(3)}$$

□ Give interpretation to the coefficients of equations (1) and (2) (no mathematical justification needed).

0.0054 is the marginal effect of disposable personal income DPI on the kitchen appliances expenditure KIT.
If DPI increases by one billion of dollars KIT increases by 0.0054 billions of dollars.
Interpretation of coefficient 0.00038 in the second equation: if DPI increases by one billion of dollars KIT increases by 0.038%.

□ What is Zarembka transformation? What is the aim of this transformation?

Zarembka transformation is $KITZ_i = KIT_i / geom.mean(KIT_i)$

It allows to compare the results of linear (in dependent variable) regression model with the logarithmic one.

□ Perform Box-Cox test to compare the models. What conclusion follows from the corresponding test? Which regression should be chosen for the further analysis?

Chi-square =10*log(0.37/0.17)=7.78. The critical value of chi-square test (1%) for 1 degree of freedom is 6.35, so the null hypotheses of no difference between equations is rejected at 1% significance level and so we can conclude that equation (3) (having less value of RSS is significantly better than the logarithmic model $\log KIT = \beta_0 + \beta_1 \cdot DPI + u$ so we choose model (1).

(b) [13 marks] Now relative price index for kitchen appliances *PRELKIT* (prices adjusted for inflation as a percentage of the 2000 level) is added to the list of regressors and the following regressors were estimated

$$\hat{KIT} = -14.2 + 8.8\log(DPI) - 8.0\log(PRELKIT) \quad RSS = 35.7 \quad R^2 = 0.97 \qquad (4)$$

$$\log(\hat{KIT}) = -10.4 + 1.4\log(DPI) - 0.3\log(PRELKIT) \quad RSS = 0.15 \quad R^2 = 0.97 \qquad \mathbf{(5)}$$

$$\hat{KITZ} = 0.96 + 0.61\log(DPI) - 0.56\log(PRELKIT) \quad RSS = 0.17 \quad R^2 = 0.97 \qquad \mathbf{(6)}$$

□ Give interpretation to the coefficients of equations (4) and (5) (no mathematical justification needed).

Coefficient 8.8 in equation (4) tells that if *DPI* increases by 1% keeping prices constant then *KIT* also increases by 0.088 billions of dollars = 88 thousands of dollars.

Coefficient –8.0 in the same equation tells that if relative prices *PRELKIT* increases by 1%. keeping income constant then *KIT* decreases by 80 thousands of dollars.

In equation (5) coefficients 1.4 and –0,3 are correspondingly income and relative prices elasticities of expenditures on kitchen appliances keeping other variable constant.

□ Perform Box-Cox test to compare the models. What conclusion follows from the corresponding test?

Chi-square =10*log(0.17/0.15)=1.25. The critical value of chi-square (5%) test for 1 degree of freedom is 3.84 , so the null hypotheses of no difference between equations is not rejected,

□ What is your choice of regression equation? How the Ramsey test can be used to justify the choice of equation? Explain.

It is possible to choose any equation (5) or (4) for the further analysis, (5) is slightly better having less value of RSS in the comparison with(6).

The Ramsey test is used to test whether some kind of nonlinearity hi the model is present. To do it for example we evaluate fitted values of $\hat{KIT}$ using equation (4) and then add $(\hat{KIT})^2$ to this equation

$$\hat{KIT} = \hat{\beta}_0 + \hat{\beta}_1 \log(DPI) + \hat{\beta}_2 \log(PRELKIT) + \hat{\beta}_3 (\hat{KIT})^2 + e_i$$

If $\hat{\beta}_3$ is significant then some kind of nonlinearity is present, maybe (5) is better. The same procedure can be applied to (5).

SECTION B.

Answer **ONE** question from this section (**3 OR 4**).

Problem 3. (25 marks) The researcher regresses the natural log of expenditure on whiskey W_t at 2020 prices in England (in GB pounds) on the natural log of total household expenditure E_t at 2020 prices (also in pounds), the natural log of the price of whiskey relative to all consumer prices PW_t (in % to 2020 level), the natural log of the price of all alcoholic drinks excluding whiskey relative to all consumer prices PA_t (also in % to 2020 level), and time t in years from $t=1$ at 2006, $t=2$ at 2007,…, till the last year in the sample 2020, with the following results

$$\underset{}{\ln W_t} = \underset{(1.387)}{-5.272} + \underset{(0.114)}{1.266}\ln E_t \underset{(0.446)}{-0.989}\ln PW_t + \underset{(0.144)}{0.412}\ln PA_t + \underset{(0.001)}{0.04}\,t + e_t$$

where e_t is the estimated residual, standard errors are in brackets, and $R^2 = 0.906$. All assumptions of the model A are assumed to be satisfied.

(a) [8 marks] □ Give brief interpretation to the coefficients in the model. Tell whether estimated values of all coefficients are reasonable from economic point of view.

Coefficients 1.266, -0.989 and 0.412 are elasticities of expenditures on whiskey correspondingly on total expenditures, whiskey prices and other alcohol prices under assumption that other variable are kept constant. The coefficient 0.004 shows that the growth rate of expenditure on whiskey is 4% per year under assumption that all other explanatory variables do not change.

All estimated values are reasonable (both in absolute values and their signs) based on economic intuition. For example price elasticities have opposite signs as increasing of whiskey prices decreases consumptions while relative increase of prices on other alcoholic drinks makes whiskey more appealing.

□ Use some math to show that your interpretations for the coefficients 1.266 (of $\ln E_t$) and 0.04 (of time t) above are correct

If $\ln W = \beta_1 + \beta_2 \ln E + \ldots$ then $\frac{\partial W}{W} = \beta_2 \frac{\partial E}{E}$ as other variables are kept constant. So $\beta_2 = \frac{\partial W}{W}\cdot 100\% : \frac{\partial E}{E}\cdot 100\%$ – elasticity (if E_t increases by one percent W_t increases by β_2 percents).

If $\ln W = \beta_1 + \ldots + \beta_5 t + \ldots$ then $\frac{\partial W}{G} = \beta_5$ keeping other variables constant. So $\frac{\Delta W}{W}\cdot 100\% = \beta_5 \cdot 100\%$, so β_5 multiplied by 100% shows the growts rate in %.

(b) [7 marks] □ Test whether coefficients of $\ln PW_t$ and $\ln PA_t$ are significant.

For $H_0 : \beta_{\ln PW_t} = 0$, $H_a : \beta_{\ln PW_t} \neq 0$, $t = \frac{-0.989}{0.446} = 2.217 < 2.228 = t_{cr}(10, 5\%)$ $\Rightarrow$ insignificant.

For $\hat{\beta}_{\ln PA_t}$, $t = \frac{0.412}{0.144} = 2.8611 > 2.228$ but $2.8611 < 3.169 = t_{cr}(10, 1\%)$ $\Rightarrow$ significant only at 5%.

□ Do your conclusions change if it can be assumed that coefficient of $\ln PW_t$ can not be positive and coefficient of $\ln PA_t$ cannot be negative?

In case of one-tailed alternative $H_0 : \beta_{\ln PW_t} = 0$, $H_a : \beta_{\ln PW_t} < 0$ we are entitled to use one-sided values 1.812 instead of 2.228 (5%) and 2.764 instead of 3.169. So now the coefficient of $\ln PW_t$ is significant and the coefficient of $\ln PA_t$ becomes significant at 1%.

(c) [10 marks] The researcher suspects that the theoretical coefficients of $\ln PW_t$ and $\ln PA_t$ are equal in absolute value being opposite in sign.

□ How to test this hypothesis against alternative hypothesis that they are not equal? What additional information is neaded for this?

For

$$\ln W_t = \beta_1 + \beta_2 \ln E_t + \beta_3 \ln PW_t + \beta_4 \ln PA_t + \beta_5 t + u_t$$

the restriction is $\beta_4 = -\beta_3$. To test this restriction we run first unrestricted equation to get R_U^2 (or RSS_U), and then to run restricted equation

$$\ln W_t = \beta_1 + \beta_2 \ln E_t + \beta_3 (\ln PW_t - \ln PA_t) + \beta_5 t + u_t$$

and then use t-test with test statistic $F = \dfrac{R_U^2 - R_R^2}{(1-R_U^2)/10}$ (or $F = \dfrac{RSS_R - RSS_U}{RSS_U/10}$). If $F > F_{crit}$ the null hypothesis $\beta_4 = -\beta_3$ is rejected in favor of alternative $H_a : \beta_4 \neq \beta_3$.

□ Is it possible to test this hypothesis against the alternative one that the coefficient of $\ln PW_t$ is greater in absolute value (being negative) than the coefficient of $\ln PA_t$ (which is positive)? What is economic meaning of these hypotheses?

Under one sided alternative $H_a : \beta_3 + \beta_3 < 0$ the F-test is useless. So we use reparametrization of the original equation

$$\ln W_t = \beta_1 + \beta_2 \ln E_t + \beta \ln PW_t + \beta_4 (\ln PA_t - \ln PW_t) + \beta_5 t + u_t \quad \text{where } \beta = \beta_3 + \beta_4$$

After running this equation we use **une sided** t-test for coefficient β: $H_0 : \beta = 0$ against $H_0 : \beta < 0$.

Aternative hypothesis $H_a : \beta_3 + \beta_3 < 0$ states that sensitivity of whiskey consumption to changes in prices of whiskey is greater (in absolute value) than its sensitivity to changes of prices of alternative alcoholic drinks.

Question 4. [25 marks] The researcher estimated (OLS) a relationship between number of students X_i enrolled a certain school and the cost of operating school Y_i for a sample of *n* schools

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

(a) [10 marks] After performing the calculation, he noticed that he forgot to include one more school in the sample, but did not recalculate.

□ Tell the researcher how TSS would change if he did include an additional school in the sample. Give reasons for your judgment.

□ How this question is connected to OLS principle applied to apropriate model? Give mathematical reasoning.

$$TSS_n = \sum_1^n (Y_i - \overline{Y}_n)^2 \overset{OLS}{\leq} \sum_1^n (Y_i - \overline{Y}_{n+1})^2 \leq \sum_1^n (Y_i - \overline{Y}_{n+1})^2 + (Y_{n+1} - \overline{Y}_{n+1})^2 = \sum_1^{n+1} (Y_i - \overline{Y}_{n+1})^2 = TSS_{n+1}$$

First inquality follows from minimization of the sum of the sum of squares (OLS): $\sum_1^n (Y_i - C)^2 \to \min$ (F.O.C.: $\dfrac{dS}{dC} = -2\sum_1^n (Y_i - C) = 0$, $nC = \sum_1^n Y_i$, $C = \dfrac{1}{n}\sum_1^n Y_i = \overline{Y}_n$) and setting $C = \overline{Y}_{n+1}$.

(b) [8 marks] In the situation of **(a)** tell the researcher if he did include an additional school in the sample

□ how RSS would change; give reasons for your judgment.

Let $\hat{Y}_i^{(n)}$ be the estimated regression values for for n sample points $(X_i, Y_i), i = 1, \ldots, n$, and $\hat{Y}_i^{(n+1)}$ be the estimated regression values for for $n+1$ sample points $(X_i, Y_i), i = 1, \ldots, n, n+1$, then

$$RSS_{n+1} = \sum_1^{n+1} (Y_i - \hat{Y}_i^{(n+1)})^2 = \sum_1^n (Y_i - \hat{Y}_i^{(n+1)})^2 + (Y_{n+1} - \hat{Y}_{n+1}^{(n+1)})^2 \geq \sum_1^n (Y_i - \hat{Y}_i^{(n+1)})^2 \overset{OLS}{\geq} \sum_1^n (Y_i - \hat{Y}_i^{(n)})^2 \geq RSS_n^*$$

So RSS also increases.

□ how R^2 would change; give reasons for your judgment.

As $R^2 = 1 - \frac{RSS}{TSS}$ and both RSS and TSS are increasing with the inclusion of the new observation, the result is unpredictable.

(c) [7 marks] The additional school differs significantly from the others in the sample: it has many more students, and the operating cost value for this school is far from the regression line plotted for the rest of the schools.

□ How do these facts affect our understanding of the validity of Gauss-Markov's conditions for the relationship in question? What are consequences of this?

A large deviation from the previous regression line for large values of the argument may indicate heterogeneity of error term (heteroscedasticity), and therefore a violation of Gauss-Markov's conditions may be assumed, which may lead to inefficient regression estimates.

□ Should a researcher simply throw this school away as an outlier? What would you suggest to do to ensure high quality of estimates?

The included additional school may have slightly different characteristics, indicating that the studied population is heterogeneous and the exclusion of additional observation will distort the modeling results. Therefore it was wrong to consider it as an outlier and simply ignore it in the sample.

To improve the modeling results it may be recommended to make the data more homogeneous, sod to consider costs per student instead of operational costs per school. We can assume that this value will be more homogeneous, which will allow at least partially eliminating heteroscedasticity.