The International College of Economics and Finance
Econometrics. Mid-term exam
2010 October 28
Marking Scheme and Solution.

Part 2. (1 hour 30 minutes). Answer the first (obligatory) question and one of the questions 2 or 3.

IMPORTANT: Start answering each question from the new page. Structure your answers in accordance with the structure of the questions. Testing hypotheses always state clearly null and alternative hypotheses provide critical value used for test, mentioning degrees of freedom and the significance level chosen for the test.

1. Answer all five parts of question 1 (8 marks each, total 40 marks)

(a) Taking each of the following statements in turn examine carefully why they might, or might not, lead to a small variance of the slope estimator in

$$Y_t = \beta_0 + \beta_1 X_t + u_t .$$

i. X values closely packed around the mean of X in your sample.
ii. X values clustered far from the origin of the X axis.
iii. Small sample sizes.
iv. Small error variance in the population regression function.

(a1) Explain your reasoning in each case.

(a2) What consequences could be drawn from here for the practice of econometrics?

(a3) Does small variance of the slope estimator mean that R^2 of regression is large? Explain.

<u>**Marking summary**</u>
Clear understanding of the distiction between population and sample characteristics [up to 2 bonus marks].
Correct formula of estimator's variance[1 mark].
Analysis of factor's influence [1 mark each].
Ability to explain consequences for practice [1 mark].
No indisputable connection between large variance of the slope estimator and R^2 .[2 marks]

a) General remarks

Remember $\operatorname{var}(\hat{\beta}_1) = \sigma^2_{\hat{\beta}_1} = \dfrac{\sigma_u^2}{\sum (X_t - \overline{X})^2}$

The students should discern the population variance and its estimator

$$\operatorname{Var}(\hat{\beta}_1) = s^2_{\hat{\beta}_1} = \frac{s_u^2}{\sum (X_t - \overline{X})^2} = \frac{\dfrac{\sum e_i^2}{n-2}}{\sum (X_t - \overline{X})^2} .$$

a1) i. $\sum (X_t - \overline{X})^2$ small $\Rightarrow$ large variance

ii. No matter how the observations are far from the origin. Only their position relative to the mean is essential. Of course it affects the precision of an intercept (it is good idea to use some graphical illustration here).

iii. The bigger the sample the bigger $\sum (X_t - \overline{X})^2$ will be and hence the smaller $\text{var}(\hat{\beta}_1) = \sigma^2_{\hat{\beta}_1} = \dfrac{\sigma_u^2}{\sum (X_t - \overline{X})^2}$ will be.

Another approach is based on the transformation of the formula with explicit indication of the impact of n $\dfrac{\sigma_u^2}{\sum (X_t - \overline{X})^2} = \dfrac{\sigma_u^2}{n\,\text{Var}(X)}$.

iv. A small σ_u^2 will imply a small $\text{var}(\hat{\beta}_1)$ (it is clear from expression for variance).

a2) To make regression more precise one should try to use as large sample as possible and pay attention to the correct specification of the model that could make the variance σ_u^2 smaller.

a3) First of all there is no connection between R^2 (sample characteristic) and population variance of the slope estimator (there is no indication on this point in UoL). Is we consider sample estimate of variance $s^2_{\hat{\beta}_1} = \dfrac{\frac{1}{n-2}\sum e_i^2}{\sum (X_t - \overline{X})^2}$ then we could discuss some links with R^2 through t-statistics and $F = \dfrac{R^2/1}{(1-R^2)/(n-k)}$, so that significance could be accompanied by larger R^2. But this link holds only 'in general', there could be regression with small R^2 and significant coefficient, so generally there is NO indisputable connection between large variance of the slope estimator and R^2.

(b) To investigate the dependence of HGC_i - highest grade completed by a respondent (in years), on his/her educational attainment $ASVABC_i$ (from 0 to 100) based on a small sample of 21 observations from the file EAEF the researcher obtained the following sample covariance matrix of these variables ($\text{Cov}(X,Y) = \dfrac{1}{n-1}\sum (X_i - \overline{X})(Y_i - \overline{Y})$; $\text{Var}(X) = \dfrac{1}{n-1}\sum (X_i - \overline{X})^2$)

	HGC	ASVABC
HGC	5.6644	14.7307
ASVABC	14.7307	92.1171

The researcher defines the new variables using demeaning of variables $h_i = HGC_i - \overline{HGC}$, $a_i = ASVABC_i - \overline{ASVCBC}$, where mean value of HGC_i is 12.95, mean value of $ASVABC_i$ is 48.55.

(b1) Find the regression $\hat{h}_i = b_1 + b_2 a_i$ and give interpretation to its coefficients.

(b2) Is the coefficient b_2 of this regression significant using 1% and 5% significance level?

Marking summary
Using formula of coefficient estimator [5 marks].
Using test [3 marks].

b1) Denote $H_i = HGC_i$, $A_i = ASVABC_i$ then slopes in regressions $\hat{H}_i = b_1^{'} + b_2^{'} A_i$ and $\hat{h}_i = b_1 + b_2 a_i$ are equal. One could find the slope from $b_1 = \dfrac{\text{Cov}(A,H)}{\text{Var}(A)} = \dfrac{14.73}{92.12} = 0.16$. Its interpretation is standard as a marginal effect of the test results (in points) on the schooling (in years). Obviously $b_1 \approx 0$ as linear regression always goes through mean point $(\overline{A}; \overline{H})$ (newertheless this is the regression with a slope!).

b2) As regression under consideration is a simple one it is possible to test its significance using F-test (all tests for simple regression are equivalent): $F = \dfrac{R^2}{(1-R^2)} \cdot (n-2)$; $R^2 = r_{a,h}^2 = \dfrac{(\text{Cov}(A,H))^2}{\text{Var}(A) \cdot \text{Var}(H)} = \dfrac{14.73^2}{92.12 \cdot 5.66} = 0.416$, so $F = \dfrac{0.416}{(1-0.416)} \cdot 19 = 13.53$. As $F_{crit.} = 8.18$ the regression is significant at 1% (and so 5%) significance level.

(c) The researcher investigates the relation between GDP – variable Y_t and money supply (M1) – variable M_t for Canada in millions of dollars for a certain period. She tries different models to fit the real data.

(c1) Explain how she would estimate the parameters α and β for the model $Y_t = \alpha M_t^{\beta}$ given data on M_t and Y_t. How the disturbance term could be installed into this equation and what assumptions need to be made about it to ensure the validity of conventional econometric tests?

(c2) The researcher runs the following regression $\widehat{\log(Y_t)} = a + bM_t$. Compare interpretation of the slope coefficient in this model with the one in (c1).

(c3) Now the researcher tries another semi-logarithmic regression $\hat{Y}_t = a + b\log M_t$. What is now the interpretation of the slope coefficient?

(c4) Is it possible to choose between different functional forms of regression in (c1) - (c3)? *Hint: first describe the comparison of regressions in (c1) with the one in (c2), then discuss the problems of comparison of regression (c3) with those in (c1) and (c2).*

Marking summary
Correct formula with disturbance term [1 mark].
Correct interpretation [1 mark each, up to 4].
Explanation of method of comparison [3 marks].

c1) We transform the equation into an estimable form by taking logs to give:

$$\log Y_t = \log(\alpha) + \beta \log(M_t),$$

i.e. a log-log model. Now we can estimate the parameters by using *OLS* — $\widehat{\log Y_t} = a + b\log(M_t)$.
Disturbance term should be included in the multiplicative form $Y_t = \alpha M_t^{\beta} v_t$ so that after taking logarithms new disturbance term would be additive $\log Y_t = \log(\alpha) + \beta \log(M_t) + \log v_t$, if $u_t = \log v_t$ has normal distribution in addition to all G-M conditions all standard tests would be valid. In this cas it is said that v_t has log-normal distribution..

c2) The interpretation of the coefficient *b* in c1) is money supply elasticity of income, that is increase in % of income as a response to the 1% increase of money supply.

In case of dependence of the form $\log(\hat{Y}_t) = a + bM_t$ the quantity 100*b shows an increase of income in % in response to the 1 billion of dollars increase of the money supply.

c3) Another semi-logarithmic regression $\hat{Y}_t = a + b\log M_t$ corresponds to the case when b/100 shows an increase of income in % in billion of dollars in response to the 1% increase of the money supply.

c4) Regressions in (c1) and (c2) with the same dependent variables are directly comparable using *RSS*, standard error of regression (choose one with smaller error term) or R^2 (choose one with greater explanatory power).

To compare regression in (c3) with dependent variable Y_t with the logarithmic regressions in (c1) and (c2) one should transform it using Zarembka transformation $Y_t^* = \dfrac{Y_t}{\sqrt[n]{Y_1 \cdot Y_2 \cdot \ldots \cdot Y_n}}$ and then to run new regression $\hat{Y}_t^* = a + b\log M_t$, that is now comparable with those in(c1) and (c2). To estimate the significance of the difference in error term *RSS* the Box-Cox test could be performed with statistics $\chi^2 = (n/2)\,|\log(RSS_{1/} / RSS_2)|$ that has chi square distribution with 1 degree of freedom.

(d) A researcher obtains data on personal expenditure on clothes, *CLOT*, disposable personal income, *DPI*, both measured in billions of dollars, and price index of clothes *PCLOT* for a developed country for the period 1976-2010. She runs a regressions using also variable TIME in years (standard errors in parenthesis),

$$\log \hat{CLOT}_t = \underset{(0.72)}{4.54} - \underset{(0.19)}{0.17} \cdot \log PCLOT_t + \underset{(0.006)}{0.044}TIME_t \qquad R^2 = 0.98 \qquad \text{(eq.1)}$$

$$\log \hat{CLOT}_t = \underset{(1.69)}{8.66} - \underset{(0.18)}{0.32} \cdot \log PCLOT_t + \underset{(0.18)}{0.46} \cdot \log DPI_t + \underset{(0.01)}{0.07}TIME_t \qquad R^2 = 0.99 \qquad \text{(eq.2)}$$

Answer some questions ignoring aspects connected with the specific of analysis of time-series models.

(d1) Give the interpretation to both equations and explain the economic meaning of their coefficients, commenting the difference in interpretation of two equations.

(d2) Comment the significance of the coefficients of these equations. Explain.

(d3) Is it reasonable to use one-sided test in order to test the significance of the coefficient of *PCLOT* in (eq.1)? Is one sided test helpful here?

(d4) First equation could be considered as a restricted version of the second equation. What is the restriction? Could it be rejected? What are in the advantages of rejection of invalid restriction and applying of valid restriction?

Marking summary

Interpretation of coefficient of TIME [1 mark].
Understanding of difference in interpretation (DPI) [2 marks].
Analysis of significance [1 mark].
One-sided tests [1 mark, extra for eq.2].
Restriction [1 mark,] advantages [1 mark each up to 2]

d1) The coefficients of log(*PCLOT*) and log(*DPI*) are correspondingly price and income elasticity of expenditures on clothes, coefficient of *TIME* shows annual rate of increase in expenditures on clothes (4.4% in eq.1) and 7% in eq.2) , all is valid under assumption that all other variables included in equation does not change.
The only difference in interpretation is the fact that in second equation when giving interpretations *DPI* should be set constant.

d2) All coefficients are significant except prices in both equations. For example the coefficient of variable log(DPI) is significant as $t = \dfrac{0.46}{0.18} = 2.55$ while $t(crit., 5\%, df = 30) = 2.042$ is less (in fact $df = (2010 - 1976 + 1) - 3 = 32$ but in case of significant result it is better to use smaller degree of freedom available from the table). On the contrary the coefficient of variable log(*PCLOT*) in second equation is insignificant as $t = \dfrac{0.32}{0.18} = 1.78$ while $t(crit., 5\%, df = 40) = 2.021$ is a little bigger (in case of insignificant result it is better to use greater degree of freedom available from the table). For the first equation $F = \dfrac{R^2/(k-1)}{(1-R^2)/(n-k)} = \dfrac{0.98/(2)}{(1-0.98)/(35-3)} = 784$ while $F(crit.,1\%, df1 = 2, df2 = 30) = 5.39$, so the null hypothesis of $H_0 : \beta(\log(PCLOT) = 0, \beta(TIME) = 0$
is rejected at 1% significance level.

d3) Using one sided test for the coefficients of log(*PCLOT*) in first equation is senseless as $|t| = \dfrac{|-0.17|}{0.19} < 1$ while $t(crit., one\ sided, 5\%, df = 30) = 1.697$ (the but in the second equation it is obviously helpful as $t = \dfrac{0.32}{0.18} = 1.78 > 1.697$ (in fact $df = (2010 - 1976 + 1) - 4 = 31$ so the rule of using smaller number of degrees of freedom is still valid).

d4) The restriction is $\beta(\log(DPI)) = 0$. It is certainly rejected as we have showed in d1) that $\beta(\log(DPI))$ is significant at 5% significance level.
Applying valid restriction reduces the risk of multicollinearity, gives additional chance to get significant results. Using invalid restriction distorts the meaning and interpretation of equation. If you are familiar with the problem of specification of regression equation some additional points could be added: using invalid restriction makes the estimates of coefficients biased and invalidates tests.

Comments: Please pay attention to the logic of the answer: with strongly minimal work all requirements to the full answer are satisfied and moreover different parts of the answer are connected. To answer in such way one should read all parts of the question and make preliminary plan of the answer.

(e) An investigator correctly believes that the relationship between two variables X and Y is given by $Y_i = \beta_1 + \beta_2 X_i + u_i$. Given a sample of observations on Y, X, and a third variable Z (which is not a determinant of Y), the investigator estimates β_2 as

$$\frac{\sum_{i=1}^{n}(Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(Z_i - \bar{Z})(X_i - \bar{X})}.$$

(e1) Demonstrate that this estimator is unbiased.

(e2) What can be said about efficiency of this estimator?

Marking summary
Derivation using expectations, clear statement and using assumptions [6 marks].
Comments on efficiency [2 marks].

e1) Noting that $Y_i - \bar{Y} = \beta_2(X_i - \bar{X}) + u_i - \bar{u}$, and taking into account that Z and X are non-stochastic

$$b_2 = \frac{\sum(Z_i - \bar{Z})(Y_i - \bar{Y})}{\sum(Z_i - \bar{Z})(X_i - \bar{X})} = \frac{\sum(Z_i - \bar{Z})\beta_2(X_i - \bar{X}) + \sum(Z_i - \bar{Z})(u_i - \bar{u})}{\sum(Z_i - \bar{Z})(X_i - \bar{X})}$$

$$= \beta_2 + \frac{\sum(Z_i - \bar{Z})(u_i - \bar{u})}{\sum(Z_i - \bar{Z})(X_i - \bar{X})}.$$

Hence as $E(u_i) = 0$ and $E(\bar{u}) = 0$ (GMC)

$$E(b_2) = \beta_2 + \frac{\sum(Z_i - \bar{Z})E(u_i - \bar{u})}{\sum(Z_i - \bar{Z})(X_i - \bar{X})} = \beta_2.$$

[6 marks]

e2) Under assumptions that G-M Conditions are satisfied OLS estimator is the most efficient.
[2 mark]

Answer ONE of TWO questions: 2 OR 3 (35 marks)

2. Suppose the model $Y_i = \beta_1 + \beta_2 X_i + u_i, \;\; i = 1, \ldots, n$ satisfies all assumption of the model A (Gauss-Markov conditions). Let the values of X_i are supposed to be non-stochastic. The researcher wrongly believes that $\beta_1 = 0$ and so uses OLS estimator $\tilde{\beta}_2$ for the model $Y_i = \beta_2 X_i + u_i, \;\; i = 1, \ldots, n$ instead of OLS estimator $\hat{\beta}_2 = \dfrac{\sum x_i y_i}{\sum x_i^2}$ for the model $Y_i = \beta_1 + \beta_2 X_i + u_i, \;\; i = 1, \ldots, n$ where $x_i = X_i - \overline{X}$, $y_i = Y_i - \overline{Y}$.

(a) Show that OLS estimator $\tilde{\beta}_2$ for the model $Y_i = \beta_2 X_i + u_i, \;\; i = 1, \ldots, n$ is $\tilde{\beta}_2 = \dfrac{\sum X_i Y_i}{\sum X_i^2}$.

Standard (see lectures) [8 marks].

a) Under assumption $\beta_1 = 0$ OLS is minimization of a function $F(\beta_2) = \sum (Y_i - \beta_2 X_i)^2$. First order condition for this is $\dfrac{d}{d\beta_2} \sum (Y_i - \beta_2 X_i)^2 = 0$ or $\sum (Y_i - \beta_2 X_i) X_i = 0$ that gives $\tilde{\beta}_2 = \dfrac{\sum X_i Y_i}{\sum X_i^2}$

(b) If the assumption $\beta_1 = 0$ is not correct show that estimator $\tilde{\beta}_2 = \dfrac{\sum X_i Y_i}{\sum X_i^2}$ is biased and find the bias.

Using formulas for expectations (GMC - assumptions of the model A should be noted [8 marks].

b) $E\tilde{\beta}_2 = E\dfrac{\sum X_i Y_i}{\sum X_i^2} = \dfrac{\sum X_i EY_i}{\sum X_i^2} = \dfrac{\sum X_i E(\beta_1 + \beta_2 X_i + u_i)}{\sum X_i^2} = \dfrac{\sum X_i (\beta_1 + \beta_2 X_i + Eu_i)}{\sum X_i^2}$. According to G.M.C. $Eu_i = 0$ so $E\tilde{\beta}_2 = \dfrac{\sum X_i (\beta_1 + \beta_2 X_i)}{\sum X_i^2} = \dfrac{\sum \beta_1 X_i + \sum \beta_2 X_i^2}{\sum X_i^2} = \beta_2 + \beta_1 \dfrac{\sum X_i}{\sum X_i^2}$. The bias is $\beta_1 \dfrac{\sum X_i}{\sum X_i^2}$.

[8 marks].

(c) From the expression of bias one could see that there is a case when $\tilde{\beta}_2$ is nevertheless unbiased, despite the fact that assumption $\beta_1 = 0$ is wrong. What is this case?

Clear explanation of the meaning of the derived conditions with possible graphical illustration [8 marks].

c) The bias could be zero in two cases: either assumption $E\beta_1 = 0$ is correct or the sample is balanced so that $\sum X_i = 0$.

(d) Compare the variances of two estimators $\tilde{\beta}_2 = \dfrac{\sum X_i Y_i}{\sum X_i^2}$ and $\hat{\beta}_2 = \dfrac{\sum x_i y_i}{\sum x_i^2}$. (Hint: variance for $\tilde{\beta}_2$ is given by expression $\text{var}(\tilde{\beta}_2) = \dfrac{\sigma^2}{\sum X_i^2}$)

Correct formula for variance (for usual OLS) [2 marks]
Clear explanation of the chosen inequality [6 marks]

d) It is known that $\text{var}(\hat{\beta}_2) = \dfrac{\sigma^2}{\sum x_i^2}$ and we are given that $\text{var}(\tilde{\beta}_2) = \dfrac{\sigma^2}{\sum X_i^2}$ ..So we need to compare both variances. Note that $\sum x_i^2 = \sum (X_i - \overline{X})^2 = \sum X_i^2 - 2\overline{X}\sum X_i + \sum (\overline{X})^2 = \sum X_i^2 - n(\overline{X})^2$ and obviously $\sum X_i^2 - n\sum(\overline{X})^2 \le \sum X_i^2$ so $\text{var}(\hat{\beta}_2) = \dfrac{\sigma^2}{\sum x_i^2} = \dfrac{\sigma^2}{\sum X_i^2 - n(\overline{X})^2} \ge \dfrac{\sigma^2}{\sum X_i^2} = \text{var}(\tilde{\beta}_2)$.

That means that if assumption $\beta_1 = 0$ is not correct $\tilde{\beta}_2$ is a superefficient estimator of a slope, but at the same time it is biased.

(e) Suppose the assumption $\beta_1 = 0$ is wrong. What estimator is better $\tilde{\beta}_2$ or $\hat{\beta}_2$ from the more general point of view, taking into account all results obtained in (a)-(c)?
(Hint: compare two estimators using the concept of $MSE(b) = \text{Var}(b) + [bias(b)]^2$).

Conditional inequality (condition in any form) [8 marks]

e) Using MSE criteria for the comparison of the estimators we get that $\tilde{\beta}_2$ is preferable to OLS estimator $\hat{\beta}_2$ $MSE(\tilde{\beta}_2) < MSE(\hat{\beta}_2)$. This is true if $\dfrac{\sigma^2}{\sum X_i^2} + \left(\beta_1 \dfrac{\sum X_i}{\sum X_i^2}\right)^2 < \dfrac{\sigma^2}{\sum X_i^2 - n(\overline{X})^2}$.

As $\sum X_i = n\overline{X}$ it could be found from here that it is equivalent to the condition

$$n(\beta_1)^2 < \sigma^2 \frac{\sum X_i^2}{\sum X_i^2 - n(\overline{X})^2}.$$

3. All students of ICEF are regularly asked to fill some (usually on-line) questionnaire grading all their teachers (more than hundred) by many criteria. The data set consists of some teacher's characteristics evaluated by students (real numbers taken as average values among students evaluations), some additional characteristics (for example teaching experience measured in years), and some dummies describing teachers (being lecturer, full time teacher with PhD, and so on). Using this data it is possible to investigate the interrelations between different criteria of teacher's success. It is supposed that the general index of teacher's success (*S*) (measured as real number from 0 to 100) correctly reflects the popularity of a teacher among the students, based on their evaluations.
The researcher runs some regressions describing the dependence of the variable *S* on some characteristics of the teachers using sample of 30 teachers.

First the researcher tries to explore the dependence of on *E* (experience of teaching at ICEF in years) and *F* (former experience of teaching in years at other Russian and foreign institutions before teaching at ICEF) and also on *T* (total teaching experience, $T = E + F$).

The results are as follows (standard errors in parenthesis)

$$\underset{(5.76)}{\hat{S} = 32.21} - \underset{(0.95)}{2.92F} + \underset{(0.83)}{2.87T} \qquad R^2 = 0.33 \qquad (3.1)$$

(a) Give interpretation (both brief and full) to all coefficients of the regression model. The researcher was worried by the unexpected sign of the coefficient of *F*, and decided that it could be explained by the insignificance of the equation or at least some parts of the equation. Do you agree with him? What is your explanation?

Correct interpretation of the constant term [1 mark]
Correct interpretation of the both coefficients with clear understanding of interaction of variables [5 marks]
No mark for any formal interpretation without special attention to the specific interdependence of variables.
Signicicance analysis with the correct conclusion [2 marks]

a) 2.87 – net marginal effect of ICEF experience (net effect implies that as former experience *F* is fixed the growth of the total experience *T* occurs entirely due to the increase of ICEF teaching *E*), minus 2.92 is a comparative disadvantage of teaching outside of ICEF (as total experience is now fixed the increase of *F* is entirely due to decreasing of ICEF teaching). (The only student of this year course that could realize and clearly explain this was Alexey Kuznetsov – congratulations!)
All parts of the equation are significant so the explanation based on multicollinearity is probably nonsense. We do not need to resort to multicollinearity to explain unusual results as it is fully clear fom the correct interpretation of coefficients.

(b) A colleague of the researcher gave him advice to use different equations: *S* as a linear function of *E* and *F*, or rather as linear function of all three variables *E*, *F* and *T*. The researcher had no time before presentation to run these regressions, but could be able to restore almost all results of the realization of the colleague' advice without computer, just using equation (3.1). Try to follow him and restore as many characteristics of these equations (coefficients, standard errors, *R*-squared) as possible. Explain your work.

Correctly restored equation [6 marks]

b) It would be a case of dummy variable trap if to include all dummies in equation (no estimates could be obtained). All coefficients of the equation with E and F as explanatory variables could be restored

$$\hat{S} = 32.21 - 2.92F + 2.87(E+F) = 32.21 + (-2.92 + 2.87)F + 2.87E = 32.21 - 0.05F + 2.87E$$

Full equation is

$$\underset{(5.76)}{\hat{S} = 32.2} - \underset{(0.95)}{0.05F} + \underset{(????)}{2.87E} \qquad R^2 = 0.33$$

As information contained in F and E is just the same as in F and T so R-squared should be the same as in the original equation, the same is true for standard errors, except the standard error of the new variable E that could not be found without running regression. There is no problem now with the negative sign of coefficient of F as this coefficient is insignificant so the data does not contradict hypothesis that the influence of former experience is also positive as the influence of ICEF teachin experience.

(c) The researcher decided to retain in equation only one factor of teaching experience E, and to take into account two additional factors L (L=1 is the teacher is a lecturer, otherwise L=0), P (P=1 if the teacher has PhD degree, otherwise L=0), and LP=L*P getting the following results (standard errors in parenthesis)

$$\underset{(5.03)}{\hat{S} = 28.54} + \underset{(0.77)}{2.27E} + \underset{(6.63)}{15.51L} + \underset{(6.95)}{5.45P} \qquad R^2 = 0.50 \quad RSS = 6495.4 \qquad (3.2)$$

$$\underset{(5.03)}{\hat{S} = 27.54} + \underset{(0.77)}{2.15E} + \underset{(8.22)}{15.54L} + \underset{(10.12)}{5.45P} - \underset{(13.51)}{18.28LP} \qquad R^2 = 0.53 \quad RSS = 6052.0 \qquad (3.3)$$

What is your conclusion about new factors taken into account in these equations? What is the difference in interpretation of equations (3.2) and (3.3). Specifically how could be interpreted the coefficient of variable LP?

Reference category should be stated, any reasonable way of explanation (graphical, algebraic, verbal) is acceptable [6 marks]

c) The best way to answer any questions on dummy variables is to specify clearly reference category. To know what is reference category simply make equal to zero all dummy variables.

Reference category in both equations (3.2) and (3.3) is inexperienced teacher giving only seminars without PhD degree. So the constant term in both equations refers to such a person (on average). According equation (3.2) the regression equation for the influence of ICEF experience E on evaluation level for non-lecturers without degree is

$$\hat{S} = 28.54 + 2.27E$$

For lecturers it becomes

$$\hat{S} = (28.54 + 15.51) + 2.27E = 44.05 + 2.27E$$

and if the lecturer additionally has PhD degree then

$$\hat{S} = (44.05 + 5.45) + 2.27E = 49.5 + 2.27E\,.$$

This equation is based on the implicit assumption that the influence of the teaching experience at ICEF is the same for all categories of teachers under consideration.

The same could be applied to the second equation, but there is an additional term with the variable LP in it that makes values of coefficients different. The coefficient of *LP* could be interpreted as an additional premium for lecturers for getting PhD degree (it is in fact negative here). Alternatively it could be interpreted as an additional (negative) premium for those who have PhD degree for being lecturers as well (this coefficient is insignificant so there is no need to pay too much attention to its negative sign).

Reference category should be stated, any reasonable way of explanation (graphical, algebraic, verbal) is acceptable [6 marks]

(d) The researcher also runs equation (3.2) separately for men and women teachers getting for men $RSS = 3619.5$ (18 observations) and for women $RSS = 443.6$ (12 observations). Is there significant difference between men and women in equation (3.2)?

Chow test is expected, main points are formula, df., sign.level, numerical result, conclusion [8 marks]

d) To evaluate the significance of the difference between the equations for men and women Chow test could be performed.

$$F = \frac{(RSS_{total} - (RSS_{male} + RSS_{female}))/k}{(RSS_{male} + RSS_{female})/n - 2k}$$

where *n* is the total number of observations and *k* is the number of parameters in each equation. So we get

$$F = \frac{(6495.4 - 3619.5 - 443.6)/4}{(3619.5 + 443.6)/22} = 3.29\,.$$

As $F(crit., 5\%, 3, 22) = 3.05$ and $F(crit., 1\%, 3, 22) = 4.82$ the difference is significant only at 5% significance level.

(e) Being interested in investigating differences between men and women in the context of the teacher's success the researcher runs the regression including in comparison with equation (3.2) also additional variables *M* (*M*=1 if the teacher is male, otherwise *M*=0), and products *ME*=*M***E*, *ML*=*M***L*, *MP*=*M***P* getting for this equation $R^2 = 0.69$. Does this support the idea of significant differences between teacher's success for men and women? What is the connection between the results obtained here and in the previous item (c)?

F-test for the group of variables is expected, main points are formula, df., sign.level, numerical result, conclusion, stating connection with the Chow test [6 marks]

d) Here we have full set of dummies so the *F*-test for this set of dummies is equivalent to the performed Chow test:

$$F = \frac{(R^2_{with\ \ male\ dummies} - R^2_{without\ \ male\ dummies})/(number\ of\ \ variables\ added)}{(1 - R^2_{with\ male\ dummies})/n - 2k}$$

$$F = \frac{(0.69 - 0.5)/4}{(1 - 0.69)/22} = 3.37$$

what is approximately equal to 3.29 in Chow test (errors due to rounding figures).

(f) How would the interpretation of equation (3.2) changes if the researcher decides to use additionally non-linear term for teaching experience at ICEF - E^2 like follows

$$\hat{S} = 33.18 - 0.07E + 0.18E^2 + 14.65L + 6.81P \qquad R^2 = 0.52$$
$$(6.84) \quad (2.47) \quad (0.18) \quad (6.69) \quad (7.08)$$

Discuss the significance of factors influencing success of a teacher taking into account that simultaneous restrictions for coefficients related with *E*: $\beta(E) = 0, \beta(E^2) = 0$ are rejected at 5% significance level.

Only explanation of the F-test for the group of variables is expected, extra marks for the formula of marginal effect [6 marks]

e) There is no direct interpretation of these coefficients, in fact they show impact of linear and nonlinear parts of the influence of the teaching experience at ICEF on the teacher's evaluation. Both are insignificant but are significant as a group (as null hypothesis is rejected at 5% significance level), so we can say that the teaching experience is significant for evaluation.
To evaluate the marginal effect of teaching experience on evaluation one could take a partial derivative of *S* with respect to the variable *E*:

$$\frac{\partial \hat{S}}{\partial E} = -0.07 + 2 \cdot 0.18E = -0.07 + 0.36E$$

so now the marginal effect is growing with the increase of teaching experience.

The International College of Economics and Finance
Econometrics. Mid-term exam. 2012 October 24
Suggested solutions

Part 2. (1 hour 30 minutes). Answer the first (obligatory) question and one of the questions 2 or 3.

IMPORTANT: Start answering each question from the new page (ask for extra paper if necessary). Structure your answers in accordance with the structure of the questions. Testing hypotheses always state clearly null and alternative hypotheses, provide critical value used for test, mentioning degrees of freedom and the significance level chosen for the test.

Answer all five parts of question 1 (8 marks each, total 40 marks)

Question 1.

(a) A researcher is studying the dependence of the total sports equipment expenditures $SPORT_t$ on the value of disposable personal income DPI_t value of taxes TAX_t, (all these three variables are measured in billions of US dollars), and relative price index of sports equipment $PRESPORT_t$ (in % of sports equipment prices to prices of total personal expenditures) for US economy in 1991-2007.

$$\underset{}{\hat{SPORT}_t} = \underset{(2.60)}{6.35} + \underset{(0.00173)}{0.0123} \cdot PI_t - \underset{(0.00652)}{0.0132} TAX_t - \underset{(0.0387)}{0.108} \cdot PRESPORT_t \qquad R^2 = 0.9707, \quad RSS = 0.9187 \qquad \text{(eq.1)}$$

(a1) Give an interpretation to the coefficient of variable TAX_t.

(a2) Test the significance of coefficients of each variable.

(a3) Test the significance of all coefficients (of all variables) simultaneously.

(a4) The researcher now assumes that the coefficients of the variables TAX_t and $PRESPORT_t$ cannot be positive and coefficient of PI_t cannot be negative. Why is the use of one-sided tests can help to find significant relationship? Does this help here?
(All questions assotiated with non-stationary time series should be ignored here).

In a2-a4 state null and alternative hypotheses, show all your work (evaluation of test statistics, comparison with critical values), explain clearly the logic of your reasoning.

(a1) If taxes increase by 1 billion of dollars the expenditures on sports equipment decrease by 13.2 millions of dollars keeping personal income and relative prices of sports equipment constant.

(a2) Let theoretical model be

$$SPORT_t = \beta_1 + \beta_2 \cdot PI_t + \beta_3 \cdot TAX_t + \beta_4 \cdot PRESPORT_t + u_t$$

To test the significance of the coefficients $\begin{pmatrix} H_0 : \beta = 0, \\ H_a : \beta \neq 0 \end{pmatrix}$ let us evaluate t-statistics for the coefficients:

$t_{PI} = \dfrac{0.0123}{0.00173} = 7.11$, $t_{TAX} = -2.03$, $t_{PRESPORT} = -2.79$ ($df = n - k = 17 - 4 = 13$),

while $t^{crit}_{5\%}(df = 13) = 2.160$, $t^{crit}_{1\%}(df = 13) = 3.012$, so coefficient of PI_t is significant at 1% level, coefficient of $PRESPORT_t$ is significant only at 5% level, and coefficient of TAX_t is insignificant.

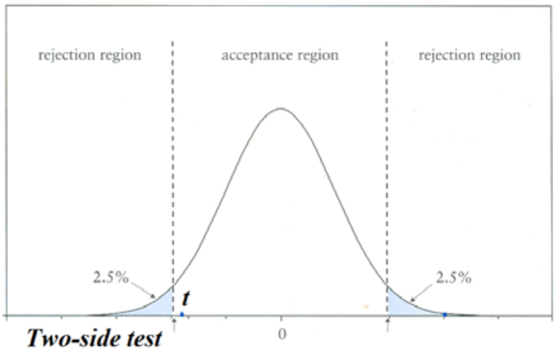
(a3) To test the significance of all coefficients simultaneously $\begin{pmatrix} H_0 : \beta_2 = \beta_3 = \beta_4, \\ H_a : \text{otherwise} \end{pmatrix}$ one should evaluate F-statistics

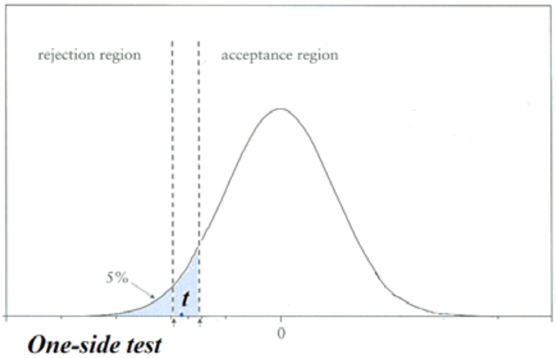
$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0.9707/3}{(1-0.9707)/13} = 143.56$$

$F^{crit}_{1\%}(3,13) = 5.74$, so H_0 is rejected.

(a4) Using one-sided tests $\begin{pmatrix} H_0 : \beta = 0, \\ H_a : \beta < 0 \end{pmatrix}$ or $\begin{pmatrix} H_0 : \beta = 0, \\ H_a : \beta > 0 \end{pmatrix}$ could help in some situations to make significant the coefficient that is insignificant when using two-sided test. For example in the situation $\begin{pmatrix} H_0 : \beta = 0, \\ H_a : \beta < 0 \end{pmatrix}$ if *t*-statistics is negative but lies in acceptance region for 5% two-sided test it could be in rejection region for one-sided test as probability to the left of the critical value is doubled now so the critical value is shifted to the right (see picture).

In our situation $t^{crit}_{5\%}(\text{one-side}, 13) = 1.771$, $t^{crit}_{1\%}(\text{one-side}, 13) = 2.65$, so coefficient of TAX_t becomes significant and coefficient of $PRESPORT_t$ becomes significant even at 1% level.





(b) *Continuation of question (a)*. The researcher decided to add new variable to her analysis - disposable personal income DPI_t, by definition $DPI_t = PI_t - TAX_t$. In addition to the former equation (eq.1)

$$\underset{}{\hat{SPORT}_t} = \underset{(2.60)}{6.35} + \underset{(0.00173)}{0.0123} \cdot PI_t - \underset{(0.00652)}{0.0132} TAX_t - \underset{(0.0387)}{0.108} \cdot PRESPORT_t \qquad R^2 = 0.9707, \; RSS = 0.9187 \qquad \text{(eq.1)}$$

she runs new regression

$$\hat{SPORT}_t = \underset{(2.60)}{6.35} + \underset{(0.00173)}{0.0123} \cdot DPI_t - \underset{(0.00517)}{0.000953} \cdot TAX_t - \underset{(0.0387)}{0.108} \cdot PRESPORT_t \qquad R^2 = 0.9707, \; RSS = 0.9187 \qquad \text{(eq.2)}$$

and then compare them

(b1) Explain why all coefficients and their standard errors except the ones of variable TAX_t are exactly the same in both equations (as well as R^2 and RSS).

(b2) Explain why the coefficient of variable TAX_t in equation (eq.2) is different from that in equation (eq.1). In fact it is less in absolute value in equation (eq.2), and insignificant, why?

A colleague advised the researcher to drop variable TAX_t from equations:

$$\hat{SPORT}_t = \underset{(2.46)}{6.44} + \underset{(0.00117)}{0.0121} \cdot DPI_t - \underset{(0.0344)}{0.108} \cdot PRESPORT_t \qquad R^2 = 0.9706, \; RSS = 0.9211 \qquad \text{(eq.3)}$$

(b3) Was this advice useful? Why do you think so? The colleague said that equation (eq.3) is a restricted version of the equation (eq.1), is it right? What is the restriction? Is it rejected?

(b1) The theoretical equation corresponding (eq.2) is

$$SPORT_t = \alpha_1 + \alpha_2 \cdot DPI_t + \alpha_3 \cdot TAX_t + \alpha_4 \cdot PRESPORT_t + u_t$$

Substituting $DPI_t = PI_t - TAX_t$ into this equation we get

$$SPORT_t = \alpha_1 + \alpha_2 \cdot PI_t - \alpha_2 \cdot TAX_t + \alpha_3 \cdot TAX_t + \alpha_4 \cdot PRESPORT_t + u_t$$

or

$$SPORT_t = \alpha_1 + \alpha_2 \cdot PI_t + (\alpha_3 - \alpha_2) \cdot TAX_t + \alpha_4 \cdot PRESPORT_t + u_t$$

what is identical to

$$SPORT_t = \beta_1 + \beta_2 \cdot PI_t + \beta_3 \cdot TAX_t + \beta_4 \cdot PRESPORT_t + u_t$$

with the same variables and the same disturbance term. So all characteristics of this equation – estimates of coefficients, their standard errors (except those related to the variable TAX_t) should be the same for both equations.

There is alternative and simpler way of explanation. It is obvious that both equation express the relationship of $SPORT_t$ from the same set of variables PI_t, TAX_t and $PRESPORT_t$. So R^2 and RSS, the coefiicient of $PRESPORT_t$ should be the same. As for coefficients of different variables PI_t and DPI_t they also should be equal, as interpretation of the coefficients of the multiple regression model requires that all other variables should remain constant: if PI_t increases by 1 dollar keeping TAX_t constant the variable DPI_t also increases by 1 dollar as it follows from equality $DPI_t = PI_t - TAX_t$ so their marginal effect to $SPORT_t$ should be the same.

(b2) The coefficient of TAX_t in equation

$$SPORT_t = \alpha_1 + \alpha_2 \cdot DPI_t + \alpha_3 \cdot TAX_t + \alpha_4 \cdot PRESPORT_t + u_t$$

is less in absolute value than in equation

$$SPORT_t = \alpha_1 + \alpha_2 \cdot PI_t + (\alpha_3 - \alpha_2) \cdot TAX_t + \alpha_4 \cdot PRESPORT_t + u_t$$

derived in (b1), taking into account that α_3 should be negative from economic consideration, see also assumptions in (**a4**). Being smaller it can easily become insignificant. Additional factor could be relatively bigger standard errors caused by multicollinearity as taxes are already included into equation through disposable personal income.

There is alternative way of explanation. The interpretation of the coefficients of the multiple regression model requires that all other variables should remain constant. In interpretation of equation

$$SPORT_t = \alpha_1 + \alpha_2 \cdot DPI_t + \alpha_3 \cdot TAX_t + \alpha_4 \cdot PRESPORT_t + u_t$$

if TAX_t increases by 1 dollar keeping DPI_t constant the variable PI_t will also increase as it follows from the equality $PI_t = DPI_t + TAX_t$. So the coefficient of TAX_t shows combined effect of simultaneous increase of TAX_t and PI_t by the same value, the result of that is economically uncertain, what is clearly explains the insignificance of this coefficient.

(b3) The equation

$$SPORT_t = \gamma_1 + \gamma_2 \cdot DPI_t + \gamma_4 \cdot PRESPORT_t + u_t$$

is the restricted version of the equation

$$SPORT_t = \alpha_1 + \alpha_2 \cdot PI_t + \alpha_3 \cdot TAX_t + \alpha_4 \cdot PRESPORT_t + u_t$$

under restriction $\alpha_3 = -\alpha_2$. It is not rejected as under $\begin{pmatrix} H_0 : \textit{no difference between equations}, \\ H_a : \textit{otherwise} \end{pmatrix}$ the F-statistics is $F = \frac{9211 - 9187}{9187} \cdot 13 = 2.01$ what is definitely insignificant ($F_{5\%}^{crit}(1,13) = 4.67$), so H_0 is not rejected, and so simplier equation (eq.3) should be chosen (more efficient estimates, less standard errors).

(c) Let the multiple linear regression equation be

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i; \quad i = 1, 2, ..., N.$$

(c1) Outline briefly, how you would test $\beta_2 = \beta_3 = 0$.

(c2) Outline briefly, how you would test $\beta_2 = \beta_3$.

(c3) Outline briefly, how you would test $\beta_2 = \beta_3$ against one-sided alternative $\beta_2 > \beta_3$

(c4) What assumptions are necessary to all of these tests were valid.

(c1) To test $\begin{pmatrix} H_0 : \beta_2 = \beta_3 = 0, \\ H_a : otherwise \end{pmatrix}$ we need to know R_1^2 from equation

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad (1).$$

Then we run standard F-test using statistics $F = \dfrac{R_1^2 / 2}{(1 - R_1^2)/(N-3)}$.

(c2) To test $\begin{pmatrix} H_0 : \beta_2 = \beta_3, \\ H_a : otherwise \end{pmatrix}$ we can run additionally restricted regression

$$Y_i = \alpha_1 + \alpha_2 (X_{2i} + X_{3i}) + u_i \quad (2)$$

and finding R_2^2 compare it with R_1^2 using F-test $F = \dfrac{(R_1^2 - R_2^2)}{(1 - R_1^2)/(N-3)}$.

Alternative way is to use standard *t*-test for coefficient $\beta_2 - \beta_3$ of the equation

$$Y_i = \beta_1 + (\beta_2 - \beta_3) X_{2i} + \beta_3 (X_{3i} + X_{2i}) + u_i \quad (3)$$

(c3) Testing $\begin{pmatrix} H_0 : \beta_2 = \beta_3, \\ H_a : \beta_2 > \beta_3 \end{pmatrix}$ is possible only on the basis of equation (3), but now one should use one-sided *t*-test.

(c4) All these tests require the disturbance term of equations (1-3) to satisfy Gauss-Markov conditions (to ensure the equations to be valid). Additionally disturbance terms should be distributed normally as it provides that statistics $t = \dfrac{\hat{\beta}}{s.e.(\hat{\beta})}$ have t-distribution for all coefficients of equations (1-3) and also the statistics $F = \dfrac{R_1^2 / 2}{(1 - R_1^2)/(N-3)}$ and $F = \dfrac{(R_1^2 - R_2^2)}{(1 - R_1^2)/(N-3)}$ have *F*-distribution with corresponding degrees of freedom.

(d) The simple linear regression model $\hat{Y}_i = \beta_1 + \beta_2 X_i + u_i$ is considered.

(d1) Demonstrate that the square of t-statistics for coefficient $\hat{\beta}_2$ is equal to the F-statistics for this equation.

The result of the estimation of the regression above using the sample of 25 observations is

$$\hat{Y}_i = 17.02 + 0.0306 \cdot X_i$$
$$(1.565) \quad (0.0086)$$

(d2) Using the result obtained in d1) find F-statistics for this equation and decide if there is significant linear relationship between X_i and Y_i.

Now new equation is obtained using another sample:

$$\hat{Y}_i = 19.36 + 0.0303 \cdot X_i \qquad R^2 = 0.3154, \qquad R^2_{adj} = 0.2665$$

(d3) Find the number of observation n and standard error of coefficient of X_i

(d1)

$$F = \frac{ESS}{RSS/(n-2)} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^{n} e_i^2 \Big/ (n-2)} = \frac{\sum_{i=1}^{n}([b_1 + b_2 X_i] - [b_1 + b_2 \bar{X}])^2}{s_u^2} =$$

$$= \frac{1}{s_u^2}\sum_{i=1}^{n} b_2^2 (X_i - \bar{X})^2 = \frac{b_2^2}{s_u^2 \Big/ \sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{b_2^2}{(\text{s.e.}(b_2))^2} = t^2.$$

(d2) Using result obtained in d1) we get $t = \dfrac{0.0306}{0.0086} = 3.56$ and then $F = t^2 = (3.56)^2 = 12.66$. It is significant at 1% level as $F_{1\%}^{crit}(1, 23) = 7.88$.

(d3) From $R^2 = 0.3154$ and $R^2_{adj} = 0.2665$ we get ,

$R^2_{adj} = R^2 - (1-R^2)\dfrac{1}{n-2}$ or $0.3154 - (1-0.3154)\dfrac{1}{n-2} = 0.2665$, so $n = 16$. Now $F = \dfrac{R^2}{(1-R^2)/14} = \dfrac{0.3154}{(1-0.3154)/14} = 6.45$, $t = \sqrt{F} = 2.54$ and finally $s.e.(b) = \dfrac{b}{t} = \dfrac{0.0303}{2.54} = 0.012$.

(e) For his coursework a student is studying the dependence of the wages W_i on the time worked T_i on the base of data on student's odd job using a sample of his friends. As job is paid only for hours actually worked the suitable model for this task is $W_i = \alpha T_i + u_i$. By ignorance he used for estimation of parameter α the standard OLS estimator $\alpha^*_{OLS} = \dfrac{\text{Cov}(W,T)}{\text{Var}(T)}$. His research advisor pointed him out that this estimator is no more valid for the model under consideration. Help the student to understand the situation (variable T_i is assumed nonstochastic, disturbance term has zero expectation, constant variance σ_u with different values not correlated each other):

(e1) Derive the OLS estimator α_{OLS} for the model $W_i = \alpha T_i + u_i$.

(e2) Show that the estimator $\alpha^*_{OLS} = \dfrac{\text{Cov}(W,T)}{\text{Var}(T)}$ is still unbiased.

(e3) Show that the variance of the estimator $\alpha^*_{OLS} = \dfrac{\text{Cov}(W,T)}{\text{Var}(T)}$ is generally speaking greater than the variance of estimator α_{OLS} for the model $W_i = \alpha T_i + u_i$ (except for some special cases).
.

(e1) To derive OLS estimator for the model $W_i = \alpha T_i + u_i$ une should minimize the sum $\sum (W_i - \alpha T_i)^2$. First order condition gives $2\sum (W_i - \alpha T_i)(-T_i) = 0$ so $\alpha_{OLS} = \dfrac{\sum W_i T_i}{\sum T_i^2}$.

(e2) Estimator $\alpha^*_{OLS} = \dfrac{\text{Cov}(W,T)}{\text{Var}(T)}$ still remains unbiased as

$E\alpha^*_{OLS} = E\dfrac{\text{Cov}(W,T)}{\text{Var}(T)} = \dfrac{\text{Cov}(EW,T)}{\text{Var}(T)} = \dfrac{\text{Cov}(E(\alpha T + u),T)}{\text{Var}(T)} = \dfrac{\text{Cov}((\alpha ET + Eu),T)}{\text{Var}(T)}$ using covariance rules. As $ET = T$ (T is non-stochastic) and $Eu = 0$ (Gauss-Markov condition) we get

$E\alpha^*_{OLS} = \dfrac{\text{Cov}((\alpha T),T)}{\text{Var}(T)} = \alpha \dfrac{\text{Cov}(T,T)}{\text{Var}(T)} = \alpha \dfrac{\text{Var}(T)}{\text{Var}(T)} = \alpha$.

(e3) As it known $\text{var}(\alpha^*_{OLS}) = \dfrac{\sigma_u^2}{\sum (T_i - \overline{T})^2}$.

On the other hand $\text{var}(\alpha_{OLS}) = \text{var}\dfrac{\sum W_i T_i}{\sum T_i^2} = \dfrac{\sum T_i^2 \,\text{var}(W_i)}{(\sum T_i^2)^2} = \dfrac{\sum T_i^2 \,\text{var}(u_i)}{(\sum T_i^2)^2} = \dfrac{\sum T_i^2 \sigma_u^2}{(\sum T_i^2)^2} = \dfrac{\sigma_u^2}{\sum T_i^2}$.

As it easily could be seen

$\sum T_i^2 = \sum ((T_i - \overline{T}) + \overline{T})^2 = \sum (T_i - \overline{T})^2 + n(\overline{T})^2 + 2\overline{T}\sum (T_i - \overline{T}) = \sum (T_i - \overline{T})^2 + n(\overline{T})^2$ as

$\sum (T_i - \overline{T}) = \sum T_i - n\overline{T} = n\overline{T} - n\overline{T} = 0$

So $\sum T_i^2 = \sum (T_i - \overline{T})^2 + n(\overline{T})^2 \geq \sum (T_i - \overline{T})^2$ hence $\text{var}(\alpha_{OLS}) = \dfrac{\sigma_u^2}{\sum T_i^2} \leq \dfrac{\sigma_u^2}{\sum (T_i - \overline{T})^2} = \text{var}(\alpha^*_{OLS})$. In case when $\overline{T} = 0$ two estimates are equally efficient.

Comment: this result could not be derived from Gauss-Markov theorem as this theorem requires intercept to be including in regression equaion.

IMPORTANT: The text in frames does not contain questions and is here just for information and your convenience. It is NOT assumed that you have to prove or comment it.

Question 2. Suppose the model $Y_i = \beta_1 + \beta_2 X_i + u_i, \quad i = 1, \dots, n$ satisfies all assumptions of the model A (Gauss-Markov conditions). The values of X_i are supposed to be non-stochastic.

As it is known the OLS estimator of β_2 is given by the expression $b_2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$.

(a) Show that b_2 is an unbiased estimator of β_2.

It is well known that $b_2 = \beta_2 + \sum_{i=1}^{n} a_i u_i$, where $a_i = \dfrac{(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$.

so $E(b_2) = E(\beta_2) + E\left\{\sum_{i=1}^{n} a_i u_i\right\} = \beta_2 + \sum_{i=1}^{n} E(a_i u_i) = \beta_2 + \sum_{i=1}^{n} a_i E(u_i) = \beta_2$ using expectation rules and $E(u_i) = 0$ (Gauss-Markov condition).

Alternative proof could be performed on the base on the formula $b_2 = \dfrac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}$ and using of the properties of expectation, variance and covariance.

(b) Show that b_2 is linear function of observed values of Y_i: $b_2 = \sum_{i=1}^{n} a_i Y_i$, where $a_i = \dfrac{(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$.

Note that

$$\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) = \sum_{i=1}^{n}(X_i - \overline{X})Y_i - \sum_{i=1}^{n}(X_i - \overline{X})\overline{Y} =$$

$$= \sum_{i=1}^{n}(X_i - \overline{X})Y_i - \overline{Y}\sum_{i=1}^{n}(X_i - \overline{X}) = \sum_{i=1}^{n}(X_i - \overline{X})Y_i - \overline{Y}\left\{\sum_{i=1}^{n} X_i - n\overline{X}\right\} = \sum_{i=1}^{n}(X_i - \overline{X})Y_i .$$

Then

$$b_2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})Y_i}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \sum_{i=1}^{n} \frac{(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} Y_i = \sum_{i=1}^{n} a_i Y_i ,$$

where coefficients a_i are defined as above.

It follows from here that the variance of the OLS estimator b_2 can be represented as $\mathrm{var}(b_2) = \sigma_u^2 \sum_{i=1}^{n} a_i^2$.

It is also well known that coefficients a_i have two properties

$$\sum_{i=1}^{n} a_i = 0 \qquad \text{(A1)}$$

and

$$\sum_{i=1}^{n} a_i X_i = 1 \qquad \text{(A2)},$$

(c) Let $\tilde{b}_2 = \sum_{i=1}^{n} g_i Y_i$ be any other linear in Y_i unbiased estimator of β_2. Using definition of unbiased estimator prove that the coefficients g_i have properties similar to the coefficients a_i:

$$\sum_{i=1}^{n} g_i = 0 \qquad \text{(G1)}$$

and

$$\sum_{i=1}^{n} g_i X_i = 1 \qquad \text{(G2)}$$

Consider any other unbiased estimator $\tilde{b}_2 = \sum_{i=1}^{n} g_i Y_i$ that is a linear function of the Y_i. For $\tilde{b}_2$ to be unbiased, we need $E(\tilde{b}_2) = \beta_2$, so

$\tilde{b}_2 = \sum_{i=1}^{n} g_i Y_i = \sum_{i=1}^{n} g_i(\beta_1 + \beta_2 X_i + u_i) = \sum_{i=1}^{n} \beta_1 g_i + \sum_{i=1}^{n} \beta_2 g_i X_i + \sum_{i=1}^{n} g_i u_i$.

Hence, $E(\tilde{b}_2) = \beta_1 \sum_{i=1}^{n} g_i + \beta_2 \sum_{i=1}^{n} g_i X_i + E\left\{\sum_{i=1}^{n} g_i u_i\right\}$. The first two terms on the right side are nonstochastic and are therefore unaffected by taking expectations. Now $E\left\{\sum_{i=1}^{n} g_i u_i\right\} = \sum_{i=1}^{n} E(g_i u_i) = \sum_{i=1}^{n} g_i E(u_i) = 0$. Thus, $E(\tilde{b}_2) = \beta_1 \sum_{i=1}^{n} g_i + \beta_2 \sum_{i=1}^{n} g_i X_i$. Hence $E(\tilde{b}_2) = \beta_2$, the g_i must satisfy $\sum_{i=1}^{n} g_i = 0$ and $\sum_{i=1}^{n} g_i X_i = 1$.

(d) Let now $g_i = a_i + b_i$. Show that properties (A1), (A2) together with (G1), G(2) implies that

$$\sum b_i = 0 \qquad \text{(B1)}$$

and

$$\sum b_i X_i = 0 \qquad \text{(B2)}$$

Now let $b_i = g_i - a_i$. Writing $g_i = a_i + b_i$, the first condition for the unbiasedness of $\tilde{b}_2$ becomes $\sum_{i=1}^{n} g_i = \sum_{i=1}^{n}(a_i + b_i) = 0$. Since $\sum a_i = 0$ (A1), this implies $\sum b_i = 0$. The second condition for the unbiasedness of $\tilde{b}_2$ becomes $\sum_{i=1}^{n} g_i X_i = \sum_{i=1}^{n}(a_i + b_i)X_i = \sum_{i=1}^{n} a_i X_i + \sum_{i=1}^{n} b_i X_i = 1$.
Since $\sum a_i X_i = 1$ (A2), this implies $\sum b_i X_i = 0$.

(e) Using representation $\tilde{b}_2 = \sum_{i=1}^{n} g_i Y_i$, where $g_i = a_i + b_i$ and taking into account properties of coefficients (B1,B2), evaluate the variance of $\tilde{b}_2$ and show that it cannot be less than the variance of OLS estimator of β_2: $\operatorname{var}(b_2) \le \operatorname{var}(\tilde{b}_2)$. Which property of estimators does this inequality mean?

The variance of $\tilde{b}_2$ is given by $\sigma^2_{\tilde{b}_2} = E\{(\tilde{b}_2 - E(\tilde{b}_2))^2\} = E\left\{\sum_{i=1}^{n}(g_i u_i)^2\right\} = \sigma_u^2 \sum_{i=1}^{n} g_i^2$.
The variance of $\tilde{b}_2$ becomes
$\sigma^2_{\tilde{b}_2} = \sigma_u^2 \sum_{i=1}^{n} g_i^2 = \sigma_u^2 \sum_{i=1}^{n}(a_i + b_i)^2 = \sigma_u^2 \left\{\sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} b_i^2 + 2\sum_{i=1}^{n} a_i b_i\right\}$.
Now $\sum_{i=1}^{n} a_i b_i = \sum_{i=1}^{n} \frac{(X_i - \overline{X})b_i}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{1}{\sum_{i=1}^{n}(X_i - \overline{X})^2}\left\{\sum_{i=1}^{n} b_i X_i - \overline{X}\sum_{i=1}^{n} b_i\right\}$.
This is zero because, as we have seen, the conditions for unbiasedness of $\tilde{b}_2$ require $\sum b_i = 0$ and $\sum b_i X_i = 0$. Hence, $\sigma^2_{\tilde{b}_2} = \sigma_u^2\left\{\sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} b_i^2\right\}$.

This must be greater than $\sigma_u^2 \sum_{i=1}^{n} a_i^2$, the variance of the OLS estimator b_2, unless $b_i = 0$ for all *i*, in which case $\tilde{b}_2$ is the same as b_2.

Question 3. A researcher has data on income *Y*, capital *K* and labor *L* indices for 1981-2007 related to the economy of some developing country (time series specific problems are out of consideration in this question). She estimated their relationship using different production functions

(a) First she runs a linear model

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

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

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

How equation (2) could be estimated using OLS? What is the meaning of this restriction (if any)?

a) Equation (1) is a linear production function, the coefficients shows partial marginal effect of capital and labor on income (standard explanations are expected). To estimate equation (2) it is sufficient to rearrange terms:
First $Y = \alpha + \theta \cdot K + L - \theta \cdot L + u$ and then $Y - L = \alpha + \theta \cdot K - \theta \cdot L + u$ which could be estimated using OLS.
The restriction is nonsense from economic point of view and has nothing to do with the condition of constant returns to scale. The sum of coefficient is close to unity simply by chance. Nevertheless imposing any restriction when it is true helps to make estimates more efficient.

(b) Then the researcher runs auxiliary regression (1a),

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

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.

b) Equation (1a) is an auxiliary equation used for testing whether the specification of the equation (1) is correct (Ramsey test). As the coefficient by the variable $(\hat{Y}_t^*)^2$ is significant one could suspect some kind of nonlinearity may be present in the data.

(c) Next she runs two logarithmic regressions

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

and

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

Give the interpretation of both models and their parameters. Write down the theoretical model (with a disturbance terms) underlying the model (4) and state assumptions necessary for its correct estimation. Prove that equation (4) is a restricted version of the equation (3). What is a restriction? Is it significant?

c) 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.
But 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 return to scale condition).

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

(d) 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:

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

NLS applied to original model (all estimates except capital elasticity are significant) :

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

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. Usually this growth rate (here 3%) is attributed to technological progress. First equation is Cobb-Douglas (CD) production function that takes into account technical 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): $Y=e^{\beta_1}K^{\beta_2}L^{\beta_3}e^{\beta_4 T}+\varpi$.

(e) Finally the researcher estimates constant elasticity of substitution (CES) function $Y = C \cdot e^{\gamma T} \cdot (u \cdot K^{-\rho} + (1-u) \cdot L^{-\rho})^{-n/\rho}$:

$$Y = 59.33 \cdot (0.87 \cdot K^{-1.72} + 0.13 \cdot L^{-1.72})^{-1/1.72} \cdot e^{0.04T} \quad R^2 = 0.97 \qquad (6)$$

Which methods could be used for estimation of CES function? What is meant by marginal rate of substitution? What is elasticity of substitution? How the estimated equations could be used for evaluation of elasticity of substitution for the Cobb-Douglas function and CES function? Comment on the rate of technological progress and the value of return to scale.

CES production function could be estimated only using NLS), as it is non-linear in parameters. It could be estimated either in original form or using preliminary transformation

$$\ln\left(\frac{Y}{L}\right) = \ln A - \left(\frac{n}{\rho}\right) \cdot \ln\left[u \cdot \left(\frac{K}{L}\right)^{-\rho} + (1-u)\right] + \gamma \cdot t$$

Marginal rate of substitution by definition is $MRS_{KL} = -\frac{dK}{dL} = \frac{Y'_L}{Y'_K}$

Elasticity of substitution by definition is $\sigma_{LK} = \frac{d\ln(K/L)}{d\ln(Y'_L/Y'_K)}$

Parameter $\rho = 1.72$ of CES function could be used for evaluation elasticity of substitution, $\sigma = \frac{1}{1+\rho}$.

We have here $\rho = 1.78$. so the elasticity of substitution for CES function is $\sigma = \frac{1}{1+1.78} = 0.37$.

In is known that CD function with constant return to scale is characterized by the elasticity of substitution equal to 1.

This equation (6) is CES production function with constant returns to scale $(n = 1)$ with technical progress. Both functions (CD and CES) here estimate the neutral technical progress with constant rate γ: $A(t) = e^{\gamma t}$.

According Cobb-Douglas function the rate of technical progress is estimated here as 3% = 0.03*100%, while CES function gives different estimate 4%.

The International College of Economics and Finance
Econometrics. Mid-term exam. 2013 October 30

Part 2. (1 hour 30 minutes). Answer the first (obligatory) question and one of the questions 2 or 3.

IMPORTANT: Start answering each question from the new page (ask for extra paper if necessary). Structure your answers in accordance with the structure of the questions. Testing hypotheses always state clearly null and alternative hypotheses, provide critical value used for test, mentioning degrees of freedom and the significance level chosen for the test.

Answer all five parts of question 1 (8 marks each, total 40 marks)

Question 1.

(a) A real estate manager is studying the dependence of the rental prices $RENT_i$ (in thousands of dollars) of flats in a big city of different factors such as size of a flat $SIZE_i$ (in square meters) and the distance from the center of the city $DIST_i$ (in hundreds of meters). First she uses data on 19 flats rented last month through her real estate agency to estimate equation

$$\hat{RENT}_i = 4.27 - 0.0091 \cdot DIST_i \qquad R^2 = 0.1874 \qquad \text{(eq.1)}$$
$$(0.399 \quad (0.0046)$$

(a1) Use confidence interval method to test the significance of the regression coefficients.

(a1) Confidence interval $b_i \pm t(crit.,95\%, df = 17) \cdot s.e.(b_1)$ is here $-0.0091 \pm 2.110 \cdot 0.0046$ or $(-0.0188; 0.0006)$. As confidence interval contains zero, the slope coefficient equal to zero is one of the acceptable hypotheses. Coefficient is insignificant.
95% confidence interval for intercept is $(3.43; 5.11)$ so the intercept is significant.

(a2) Help the manager to get full interpretation of the obtained equation.

(a2) Coefficient 0.0091 of $DIST_i$ shows marginal effect of the distance from the center of the city on the rental prices: if distance increases by 1 hundred meters the rental prices drops by 9.1 dollars on average.
The estimate of the rental price of typical flat near the center of the city is 4.27 thousands of dollars.

This interpretation of the intercept is valid if the sample is really contains data on rented flats near the center. It should be noted that since the constant equation is significant, and the coefficient is not significant, the equation can be interpreted as an estimate of the average cost of renting an apartment in the city center.

(a3) Alternative method to confidence intervals is significance testing. Help her to test the significance of the coefficients. Take into account that she was not sure in the signs of coefficients in advance before the regression was estimated.

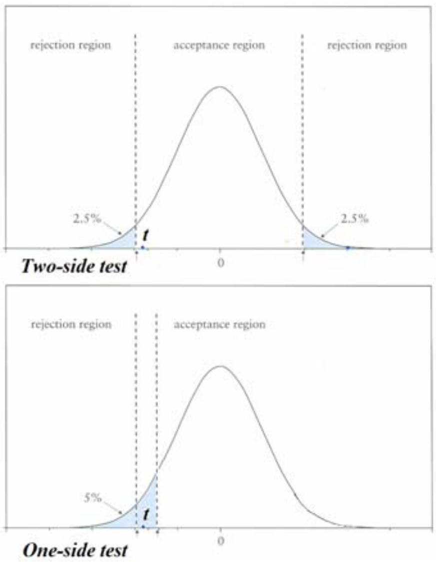
(a3) As the researcher is not sure in the sign of coefficients only two-sided test is acceptable here. Let theoretical model be

$$RENT_t = \beta_1 + \beta_2 \cdot DIST_t + u_t$$

To test the significance of the slope coefficients $\begin{pmatrix} H_0 : \beta_i = 0, \\ H_a : \beta_i \neq 0 \end{pmatrix} i = 1, 2$ let us evaluate t-statistics for the coefficients: $t_{const} = \dfrac{4.27}{0.399} = 10.70$, $t_{DIST} = -1.978$, $t(crit.,5\%,2-sided, df = 17) = 2.110$, $t(crit.,1\%,2-sided, df = 17) = 2.898$ As $|-1.978| < 2.110$ the slope coefficient is insignificant, while intercept is significant even at 1% level.

(a4) Explain the researcher why having founded assumptions about the signs of regression coefficients to calculate, one can, in some cases justify their significance. Does it really help here? Explain how to use this method correctly here stating clearly the null and alternative hypotheses.

(a4) Using one-sided tests $\begin{pmatrix} H_0 : \beta = 0, \\ H_a : \beta < 0 \end{pmatrix}$ or $\begin{pmatrix} H_0 : \beta = 0, \\ H_a : \beta > 0 \end{pmatrix}$ could help in some situations to make significant the coefficient that is insignificant when using two-sided test. For example in the situation $\begin{pmatrix} H_0 : \beta = 0, \\ H_a : \beta < 0 \end{pmatrix}$ if t-statistics is negative but is in acceptance region for 5% two-sided test it could be in rejection region for one-sided test as probability to the left of the critical value is doubled now so the critical value is shifted to the right (see picture).



In our situation $t_{5\%}^{crit}(\text{one}-\text{side}, 17) = 1.740$, so coefficient of $DIST_t$ becomes significant 5% level.

(b) *Continuation of question (a)*. The researcher noticed that usually flats that are located far from the center are bigger in size. Taking into account both factors simultaneously she gets equation (eq.2)

$$\widehat{RENT}_i = \underset{(0.3844)}{4.4800} - \underset{(0.0062)}{0.0180} \cdot DIST_i + \underset{(0.0031)}{0.0061} \cdot SIZE_i \qquad R^2 = 0.3525 \qquad \text{(eq.2)}$$

and then compares equations (eq.1) and (eq.2). Correlation between variables $DIST_i$ and $SIZE_i$ is 0.73.

(b1) Give again the full interpretation to all coefficients of this equation and explain based on this interpretation why coefficient of variable $DIST_i$ is now greater in absolute value than in equation (eq.1). Could you predict the estimates in (eq.1) having the ones in (eq.2)?.

(b1) Now coefficient of $DIST_i$ shows the marginal effect of distance when $SIZE_i$ is fixed, while coefficient of $SIZE_i$ shows the marginal effect of flat's size on the renral price when distancefrom the center is fixed. This means for example that the rental price drops by 180 dollars on average if distance from the center rises by hundred of dollars keeping size of the flat fixed.
Similarly for the coefficient of the size of apartment. The intercept now formally has no interpretation as there can be no apartment of zero size. On the other hand we can see that the value of the intercept remained nearly the same as in (eq.1). This can be explained by the fact that coefficient of $SIZE_i$ is

insignificantly differs from zero so intercept can be interpreted in the same way as in equation (eq.q) that is the estimate of a rental price for the typical flat in the center of the city.

The coefficient of the $SIZE_i$ is now greater and it is now significant, that can be explained based on the interpretation. It is said that variables $DIST_i$ and $SIZE_i$ in reality usually move together in the same direction affecting prices in the opposite way. The result of a compromise between the opposing influences is the slope coefficient in equation (eq.1). The coefficient of the $DIST_i$ in (eq.2) shows net effect of distance on the prices ('net' means that size is fixed).
Another explanation can be given based on the theory of specification of the multiple regression
Let theoretical equation corresponding (eq2) is

$$RENT_t = \beta_1 + \beta_2 \cdot DIST_i + \beta_3 \cdot SIZE_i + u_i$$

and suppose this specification correct.
If the variable $SIZE_i$ is excluded from equation the coefficient of $DIST_i$ will be estimated with the bias equal to

$$bias(\hat{\beta}_2) = \beta_3 \cdot \frac{\mathrm{Cov}(DIST_i, SIZE_i)}{\mathrm{Var}(DIST_i)}$$

The theoretical coefficient of size β_3 can be supposed to be positive (it is positive in fact in equation 2, supposed to be correct) and the sign of covariance coincides with the sign of correlation (that is equal to 0.73), so the bias should be positive. This is the just the situation that we observe in equation (eq.1).
For exact prediction the values of $\mathrm{Cov}(DIST_i, SIZE_i)$ and $\mathrm{Var}(DIST_i)$ in covariance matrix should be known.

(b2) Equation (eq.1) is a restricted version of equation (eq.2). What is the restriction? Is it significant?

(b2) If theoretical equation corresponding (eq2) is

$$RENT_t = \beta_1 + \beta_2 \cdot DIST_i + \beta_3 \cdot SIZE_i + u_i$$

then the restriction is

$$\beta_3 = 0$$

It is insignificant as $t_{SIZE} = \frac{0.0061}{0.0031} = 1.968$, while $t(crit., 5\%, df = 16) = 2.120$.

The alternative way of testing restriction is using F-test: $F = \frac{R^2_{unrestricted} - R^2_{restricted}}{(1 - R^2_{unrestricted}) / df\,(unrestricted)}$ we have

$F = \frac{0.3525 - 0.1874}{(1 - 0.3525)/16} = 4.08$ while $F(crit., 5\%, df = 1, 16) = 4.49$.

(b3) Are two variables $DIST_i$ and $SIZE_i$ taken together significant.

(b3) Null hypothesis for epy equation

$$RENT_t = \beta_1 + \beta_2 \cdot DIST_i + \beta_3 \cdot SIZE_i + u_i$$

is $H_o : \beta_2 = 0, ; \beta_3 = 0$

$F = \frac{R^2/2}{(1 - R^2)/df} = \frac{0.3525}{(1 - 0.3525)/16} = 4.355$ while $F(crit., 5\%, df = 2, 16) = 3.63$, so eauation as a whole is significant

(b4) Explain why the standard error of the coefficient of the variable $DIST_i$ in (eq.2) is larger than in (eq.1). Specify the other shortcomings of the equation (2), and discuss their probable causes.

(b4) The theoretical value for the standard error of the coefficient of $DIST_i$ in (eq.1) is $\text{s.e.}(b_2)=\sqrt{\dfrac{s_u^2}{\sum(X_{2i}-\bar{X}_2)^2}}$ while the one for equation (eq.2) is $\text{s.e.}(b_2)=\sqrt{\dfrac{s_u^2}{\sum(X_{2i}-\bar{X}_2)^2}\times\dfrac{1}{1-r^2_{X_2,X_3}}}$.

As $r^2_{X_2,X_3}=0.73^2$ the factor is $\text{s.e.}(b_2)=\sqrt{\dfrac{1}{1-r^2_{X_2,X_3}}}=\sqrt{\dfrac{1}{1-0.73^2}}=1.46$ when actual ratio is $\dfrac{0.062}{0.046}=1.35$.

The observed situation when equation as a whole is significant while one of the variables ($SIZE_t$) is insignificant indicates on the presence of multicollinearity caused probably by the high correlation between $DIST_i$ and $SIZE_t$ equal to 0.73.

(c) The rise in prices for public transport leads to lower corporate earnings, as people tend to choose cheaper alternatives. The student tries to find the best form of dependence of the volume of transportation T_i of some 50 transportation companies (in millions of dollars) from the prices of transportation P_i (in cents per one kilometer of transportation). She runs regressions (1-4) (linear, logarithmic and semi-logarithmic functions), she also runs two auxiliary regressions (5-6) performing Zarembka transformation (variable TZ_i is defined as $TZ_i=T_i/\sqrt[n]{T_1\cdot T_2\cdot\ldots\cdot T_n}$):

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Dependent variable	T_i	T_i	$\log(T_i)$	$\log(T_i)$	TZ_i	TZ_i
Independent variable\Constant	8.74	12.26	2.175	2.635	1.171	1.641
P_i	-0.339	-	-0.0045		-0.0045	
$\log(P_i)$	-	-1.362	-	-0.179	-	-0.179
R^2	0.638	0.738	0.665	0.755	0.638	0.738
RSS	4.481	3.247	0.068	0.051	0.080	0.058

(c1) Explain the differences in the values of a slope coefficient in regression (1) and (4) giving interpretation to both regressions.

(c1) The equations (1) and (4) have different specifications: (1) is linear while (4) is double-logarithmic, so the interpretation of their slope coefficients are different

For (1) coefficient of P_i equal to -0.339 shows the marginal effect of prices on the volume of transportation, namely if the prices rise by 1 cent per one kilometer the volume of transportation drops by 339 thousands of dollars.

For (4) coefficient of $\log(P_i)$ equal to -0.179 shows the price elasticity of the volume of transportation, namely if the price per one kilometer of transportation rises by 1 percent the volume of transportation drops by 0.179 percent. They are different as they have different economic meaning, but they both are negative expressing negative influence of the rising of prices on the volume of transportation.

(c2) Explain the differences in the values of a slope coefficient in regression (2) and (3) giving interpretation to both regressions.

(c2) The equations (2) and (3) are both semi-logarithmic regressions but of different specifications: Equation (2) is linear-logarithmic regression while (3) is log-linear regression, so the interpretation of their slope coefficients are different.

For (2) coefficient of $\log(P_i)$ equal to -1.362 shows that if the prices rise by 1% the volume of transportation drops by $\frac{-1.362}{100} = -0.01362$ millions of dollars or by 13.62 thousands of dollars on average.

For (3) coefficient of P_i equal to -0.0045 shows that if the price of transportation rises by 1 cent per one kilometer the volume of transportation drops by $-0.0045 \cdot 100 = -0.45$ percent on average.

(c3) Explain using some math why your interpretation is correct (for one of the regressions 2-4).

(c3) It is sufficient here to present one of the following four sets of mathematical reasoning.

For model (4). Direct method.

Let the dependence of volume of transportation of prices has specification

$Y = \beta_1 X^{\beta_2}$. Then

$$\frac{dY}{dX} = \beta_1 \beta_2 X^{\beta_2 - 1}$$

On the other hand

$$\frac{Y}{X} = \frac{\beta_1 X^{\beta_2}}{X} = \beta_1 X^{\beta_2 - 1}$$

So elasticity is $\frac{dY/dX}{Y/X} = \frac{\beta_1 \beta_2 X^{\beta_2 - 1}}{\beta_1 X^{\beta_2 - 1}} = \beta_2$

OR

For model (4). Method of differentials.

Taking logarithms of both sides of $Y = \beta_1 X^{\beta_2}$ we get the function of the form $\log Y = a + b \cdot \log X$ then

$\frac{dY}{Y} = d(a + b \cdot \log X) = b \cdot \frac{dX}{X}$, from here

$$b = \frac{\frac{dY}{Y}}{\frac{dX}{X}} = \frac{dY}{dX} \cdot \frac{X}{Y} = \frac{\frac{dY}{Y} \cdot 100\%}{\frac{dX}{X} \cdot 100\%}$$

setting $\frac{dX}{X} \cdot 100\% = 1\%$ obtain $b = \frac{dY}{Y} \cdot 100\%$

so b shows the percentage increase of Y when X increases by one percent.

OR

For model (2). Method of differentials.

Let $Y = a + b \cdot \log X$ then

$dY = d(a + b \cdot \log X) = b \cdot \dfrac{dX}{X}$, from here

$b = \dfrac{dY}{\frac{dX}{X}}$ or $\dfrac{b}{100} = \dfrac{dY}{\frac{dX}{X} * 100(\%)}$

setting $\dfrac{dX}{X} \cdot 100\% = 1\%$ obtain $\dfrac{b}{100} = dY$

If you divide *b* by 100, the resulting number will show the increase of *Y*, provided *X* will increase by one percent (the units of measurement should always be clearly indicated)

OR

For model (3). Method of differentials.

Let $\log Y = a + bX$ then

$\dfrac{dY}{Y} = d(a + bX) = bdX$, from here

$\dfrac{dY}{Y} \cdot 100\% = b \cdot 100\% dX$ or $b \cdot 100\% = \dfrac{\frac{dY}{Y} \cdot 100\%}{dX}$

setting $dX = 1$ obtain $b \cdot 100\% = \dfrac{dY}{Y} \cdot 100\%$

If you multiply *b* by 100, the resulting number will show the percentage increase of *Y*, provided *X* will increase by one unit (the units of measurement should always be clearly indicated)

Other ways of explaining how to interpret the coefficients are also welcomed. (the students who did this hey were awarded by bonus marks).

(c4) Which pairs of regression are comparable directly without Zarembka transformation). Which regressions becomes comparable after Zarembka transformation? Compare some regressions performing appropriate tests.

(c4) The regressions (1) and (2) and regressions (3) and (4) are comparable directly, as in these pairs the dependent variables are the same and so their TSS's (Total Sums of Squares) are the same. So the comparison can be made on the basis of their RSS's or determination coefficients: (2) is better than (1) and (3) is better than (4) (we do not discuss here the significance of the differences).

To compare (1) and (3) and to compare (2) and (4), and also to estimate the significance of the difference in their quality one need to perform Box-Cox test on the base of the auxiliary regressions (5) and (6) correspondingly that uses Zarembka transformation of the dependent variable.

As it is known the value $\chi^2 = (\frac{n}{2})\left|\log\left(\frac{RSS1}{RSS2}\right)\right|$ has χ^2-distibution with 1 degrees of freedom.

For the comparison of (1) and (3) (and so using (3) and (5)) we get

$\chi^2 = (\frac{50}{2})\left|\log\left(\frac{0.080}{0.68}\right)\right| = 4.063$ what is greater than 5% critical value of chi-squared for 1 degree of freedom 3.84. So the diffefence in quality of regressions is significant and the student shoud choose the regression (3) and drop out the regression (1) as having less quality.

The same test in pair (2) and (4) (using for the test (4) and (6)) gives

$\chi^2 = (\frac{50}{2})\left|\log\left(\frac{0.058}{0.051}\right)\right| = 3.215$ what is less than 5% critical value of chi-squared for 1 degree of freedom 3.84. So the diffefence in quality of regressions is insignificant and the student can choose any of the regressions (2) or (4). For example, she could choose (4) as it has clear economic interpretation using elasticity and also has less RSS.

(d) The researcher A is studying hourly earnings of freelancers working on a fashion magazine. She asked ten freelancers taken randomly from the list of all employees in this category to tell her the number of hours required to write a material for a fashion magazine, H_i, and the amount of money received as payment, P_i , in hundreds of dollars (the publishing house does not pay freelancers for the number of hours worked, but for the quantity and quality of the material).

Being inexperienced in econometrics she simply divided the payment for each material by the hours spent to prepare this material (variable $R_i = P_i / H_i$ - ratio)), and then used the average ratio $\overline{R} = \frac{\sum R_i}{10}$ as the regression coefficient of the model

$$\hat{P}_i = 0.9008 H_i \qquad \text{(eq. A)}$$

Her friend B also inexperienced in econometrics said her to find the regression line it is better to divide mean value of payment by the mean value of hours worked so the regression is

$$\hat{P}_i = 0.8571 H_i \qquad \text{(eq. B)}$$

Their colleague C a little more inexperienced in econometrics said them that they are both wrong and correct way to find the regression line is to use computer and to run one of regression by OLS

$$\underset{}{\hat{P}_i} = \underset{(0.74)}{-0.64} + \underset{(0.18)}{1.039} H_i \quad R^2 = 0.8 \qquad \text{(eq. C1)}$$

or

$$\hat{P}_i = \underset{(0.089)}{0.9006} H_i \quad R^2 = 0.78 \qquad \text{(eq. C2)}$$

A and B objected to him saying that equation (eq. C1) is incorrect both from economic point of view and econometrically and overestimates the coefficient of the variable H_i.

(d1) Comment the discussion of the researcher A and B and their friend C. What specification of the model is more appropriate here, the model with intercept or the model without intercept and why?

(d1) The question relates to economic meaning of the data under consideration. It is said clearly that the publishing house pays to freelancers only for the work done actually, and so there is no fixed part in their payments. So the regression without intercept perfectly fits the situation. Three regressions (eq.A), (eq.B) and (eq.C2) are of this type. On the other hand there can be some economic arguments in favor of inclusion of the intercept into equation.

(d2) Are A and B right saying that equation (eq.C1) is incorrect econometrically and the equation in this specification cannot be used for the estimation of regression coefficient in the problem under consideration? Comment their remark that this equation overestimates the coefficient of the variable H_i.

(d2) Now let us discuss the econometric consequences of the estimation of the equation specified as

$$P_i = \beta_1 + \beta_2 H_i + u_i$$

under assumption that correct specification is

$$P_i = \beta_2 H_i + u_i$$

First let us show that the estimate of β_2 remains unbiased. Suppose all Gauss-Markov conditions are satisfied and H_i are supposed to be non-stochastic.

Estimator of β_2 using OLS applied to the usual linear regression with intercept $P_i = \beta_1 + \beta_2 H_i + u_i$ is $\hat{\beta}_{OLS} = \dfrac{\mathrm{Cov}(P,H)}{\mathrm{Var}(H)}$.

There are several ways to prove that $\hat{\beta}_2$ is still unbiased.

The simplest way is to use the formula through covariance and variance. Substituting $P_i = \beta_2 H_i + u_i$ into covariance in the expression for OLS estimator $\hat{\beta}_{OLS} = \dfrac{\mathrm{Cov}(P,H)}{\mathrm{Var}(H)}$

$$\hat{\beta}_2 = \frac{\mathrm{Cov}(P,H)}{\mathrm{Var}(H)} = \frac{\mathrm{Cov}(\beta_2 H + u, H)}{\mathrm{Var}(H)} = \frac{\beta_2 \mathrm{Cov}(H,H) + \mathrm{Cov}(u,H)}{\mathrm{Var}(H)} = \beta_2 + \frac{\mathrm{Cov}(u,H)}{\mathrm{Var}(H)}$$

As $\mathrm{Cov}(H,H) = \mathrm{Var}(H)$ and $\mathrm{Cov}(\beta_1, H) = 0$. So

$$E\hat{\beta}_2 = E\beta_2 + E\frac{\mathrm{Cov}(u,H)}{\mathrm{Var}(H)} = \beta_2 + \frac{\mathrm{Cov}(Eu,H)}{\mathrm{Var}(H)} = \beta_2 + \frac{\mathrm{Cov}(0,H)}{\mathrm{Var}(H)} = \beta_2.$$

COMMENT: Another good proof is based on the well known decomposition $\hat{\beta}_2 = \beta + \sum a_i u_i$ where $a_i = \dfrac{\sum (H_i - \overline{H})}{\sum (H_i - \overline{H})^2}$. The proof itself is very simple $E\hat{\beta}_2 = E\beta + E(\sum a_i u_i) = \beta + (\sum a_i E u_i) = \beta$ as $Eu_i = 0$ (G-M condition).

But this decomposition was obtained under assumption that the true equation is $P_i = \beta_1 + \beta_2 H_i + u_i$. So this property (decomposition) should be proved first under new assumptions.

But unbiasedness is not the only criterion for the characterizing quality of the estimators.

When using equation with intercept the estimation of one extra parameter β_1 leads to increasing of variances of estimators of both coefficients so they will be not the most efficient, the standard errors of its coefficients becomes generally speaking greater.

It could be clearly seen using comparison of the coefficients of two equations

$$\underset{(0.74)}{\hat{P}_i = -0.64} + \underset{(0.18)}{1.039} H_i \quad R^2 = 0.8 \qquad \text{(eq. C1)}$$

$$\hat{P}_i = \underset{(0.089)}{0.9006} H_i \quad R^2 = 0.78 \qquad \text{(eq. C2)}$$

The standard error of the slope coefficient 0.18 in (eq. C1) is twice as big as the standard error of the slope coefficient 0.089 in (eq. C2).

So the relatively large value of the slope coefficient 1.039 in (eq.C1) can be explained not by its systematic bias upwards (it is unbiased!) but rather by some random factors as under larger standard error the larger deviations are quite probable. **Comment**: As intercept appears to be negative the slope should be bigger to fit the observations (Andrey Telegin, Econometrics 2013-14).

(d3) C replied saying that the equation (eq. C2) is superior to equations (eq. A) and (eq. B) as the Gauss-Markov theorem ensures that the variance of the coefficient of (eq. C2) is minimal. Comment on this remark of C.

(d3) One of the Gauss-Markov Conditions (or assumptions of model A) is the assumption that the linear model contains intercept (it is explicitly present in the outline of the proof both in the textbook and the lecture. So the reference to the Gauss-Markov theorem is incorrect.
Comment: On the other hand presumably it is possible to prove that OLS estimator is still the best even in less popular in practice case of regression without intercept (any reasonable suggestion of this sort gives bonus mark).

(e) The simple linear regression without intercept is considered where X_i are non-stochastic:

$$Y_i = \beta X_i + u_i;\ Eu_i = 0,\ \text{var}(u_i) = \sigma;\ i = 1, 2, ..., n$$

Two estimators of β are suggested: $\hat{\beta}_1 = \overline{Y} / \overline{X}$ and $\hat{\beta}_2 = \overline{Y_i / X_i}$

(e1) Which one (if any) is OLS estimator of β ?

(e1) OLS estimator for model under consideration is $\hat{\beta} = \dfrac{\sum Y_i X_i}{\sum X_i^2}$ so it is neither $\hat{\beta}_1 = \overline{Y} / \overline{X}$ nor $\hat{\beta}_2 = \overline{Y_i / X_i}$. It could be derived from the OLS principle

$$\sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - \hat{\beta}_{OLS} X_i)^2 \to \min$$

Taking the derivative with respect to $\hat{\beta}_{OLS}$ we get

$$-2\sum X_i (Y_i - \hat{\beta}_{OLS} X_i) = \sum X_i Y_i - \hat{\beta}_{OLS} \sum X_i^2 = 0$$

From here $\hat{\beta} = \dfrac{\sum Y_i X_i}{\sum X_i^2}$

A good answer might go on to show that the stationary point where the derivative =0 is a minimum: let S be the sum of residuals. Let us take the second derivative

$\dfrac{\partial^2 S}{\partial \hat{\beta}^2} = \dfrac{\partial}{\partial \hat{\beta}} \left(-2\sum_{t=1}^{T} (Y_t X_t - \hat{\beta} X_t^2) \right) = 2\sum_{t=1}^{T} X_t^2 > 0$ what indicated that here we have point of minima.

(e2) Prove that both $\hat{\beta}_1$ and $\hat{\beta}_2$ are unbiased estimators of β.

(e2) $$\hat{\beta}_1 = \frac{\overline{Y}}{\overline{X}} = \frac{\frac{1}{n}\sum Y_i}{\frac{1}{n}\sum X_i} = \frac{\sum(\beta X_i + u_i)}{\sum X_i} = \beta + \frac{\sum u_i}{\sum X_i}$$

hence $E(\hat{\beta}_1) = \beta + \dfrac{\sum E(u_i)}{\sum X_i} = \beta$ if $E(u_i) = 0$.

$$E(\hat{\beta}_2) = E(\frac{1}{n}\sum(Y_i / X_i)) = \frac{1}{n}E(\sum(Y_i / X_i)) = \frac{1}{n}\sum\frac{E(Y_i)}{X_i} = \frac{1}{n}\sum\frac{\beta X_i}{X_i} = \frac{1}{n}\sum\beta = \frac{n\beta}{b} = \beta$$

Hence $\hat{\beta}_2 = \overline{Y_t / X_t}$ is an unbiased estimator.

In both proofs the assumptions that $E(u_i) = 0$ and X_t is non-stochastic were used.

(e3) Outline briefly how you would choose between two estimators $\hat{\beta}_1$ and $\hat{\beta}_2$.

(e3) As both estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ are unbiased we can choose the most efficient estimator. The unbiased estimator is called efficient relatively to another unbiased estimator if its theoretical variance is relatively less than the variance of another estimator. So to choose between estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ it is sufficient to find their variances and compare them for any sets of data.

Comment: it was not expected that student can derive the expressions for the variances of the estimators under consideration and then compare them, but some students did this (they were awarded by bonus marks).

ANSWER ONE OF TWO QUESTIONS: 2 OR 3 (total 35 marks)

Question 2. Suppose the model $Y_i = \beta_1 + \beta_2 X_i + u_i$, $i = 1, \dots, n$ satisfies all assumptions of the model A (Gauss-Markov conditions). The values of X_i are supposed to be non-stochastic.

(a) Show that $b_1 = \beta_1 + \sum_{i=1}^{n} c_i u_i$, where $c_i = \frac{1}{n} - a_i \overline{X}$ where $a_i = \frac{X_i - \overline{X}}{\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2}$.

$b_1 = \overline{Y} - b_2\overline{X} = (\beta_1 + \beta_2\overline{X} + \overline{u}) - \overline{X}(\beta_2 + \sum a_i u_i) =$

$= \beta_1 + \frac{1}{n}\sum u_i - \overline{X}\sum a_i u_i = \beta_1 + \sum(\frac{1}{n} - a_i\overline{X})u_i = \beta_1 + \sum c_i u_i$, where $c_1 = \frac{1}{n} - a_i\overline{X}$

(b) Investigate the properties of coefficients c_i. Prove that

$\sum_{i=1}^{n} c_i = 1$, **(C1)**

$\sum_{i=1}^{n} c_i X_i = 0$. **(C2)**

$\sum c_i = \sum(\frac{1}{n} - a_i\overline{X}) = \sum(\frac{1}{n}) - \overline{X}\sum a_i = 1 - \overline{X}\cdot 0 = 1$

$\sum c_i X_i = \sum(\frac{1}{n} - a_i\overline{X})X_i = \sum(\frac{X_i}{n}) - \overline{X}\sum a_i X_i = \overline{X} - \overline{X} = 0$

(c) Show that OLS estimator of the intercept b_1 is linear unbiased estimator of β_1.

Part one: b_1 is unbiased.

Using the decomposition $b_1 = \beta_1 + \sum_{i=1}^{n} c_i u_i$ where we get

$E(b_1) = E\left(\beta_1 + \sum c_i u_i\right) = \beta_1 + \sum c_i E(u_i) = \beta_1$

Part two: b_1 is linear function of Y_i.

We know that $b_2 = \sum_{i=1}^{n} a_i Y_i$, where $a_i = \frac{(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$ so.

$b_1 = \overline{Y} - b_2\overline{X} = \frac{1}{n}\sum_{i=1}^{n} Y_i - \overline{X}\sum_{i=1}^{n} a_i Y_i = \sum_{i=1}^{n}(\frac{1}{n} - a_i\overline{X})Y_i = \sum_{i=1}^{n} c_i Y_i$

where c_i are defined as above $c_1 = \frac{1}{n} - a_i\overline{X}$.

(d) Prove that the variance of OLS estimator of the intercept is

$$\sigma_{b_1}^2 = \sigma_u^2 \sum c_i^2 = \sigma_u^2\left\{\frac{1}{n} + \frac{\overline{X}^2}{\sum\left(X_i - \overline{X}\right)^2}\right\}$$

$b_1 = \beta_1 + \sum c_i u_i$, where $c_1 = \frac{1}{n} - a_i \overline{X}$. Hence

$$\sigma_{b_1}^2 = E\left[\left(\sum c_i u_i\right)^2\right] = \sigma_u^2 \sum c_i^2 = \sigma_u^2 \left(n\frac{1}{n^2} - 2\frac{\overline{X}}{n}\sum a_i + \overline{X}^2 \sum a_i^2 \right).$$

As it is known, $\sum a_i = 0$ and $\sum a_i^2 = \frac{1}{\sum \left(X_i - \overline{X}\right)^2}$.

Hence $\sigma_{b_1}^2 = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum \left(X_i - \overline{X}\right)^2} \right\}$.

The alternative way to prove this is to use directly properties of variance

It should be mentioned that if u_i are uncorrelated then Y_i are also uncorrelated

Using the decomposition $b_1 = \beta_1 + \sum_{i=1}^{n} c_i u_i$ where $c_1 = \frac{1}{n} - a_i \overline{X}$ and $a_i = \frac{X_i - \overline{X}}{\sum_{j=1}^{n} \left(X_j - \overline{X}\right)^2}$ we get

$$\operatorname{var}(b_1) = \operatorname{var}(\beta_1 + \sum_{i=1}^{n} (\frac{1}{n} - a_i \overline{X}) u_i) = \operatorname{var}(\sum_{i=1}^{n} (\frac{1}{n} - a_i \overline{X}) u_i) = \sum_{i=1}^{n} (\frac{1}{n} - a_i \overline{X})^2 \operatorname{var}(u_i)$$

(as u_i are uncorrelated according to assumptions of model A or G-M Conditions)

$$= \sigma_u^2 \sum_{i=1}^{n} (\frac{1}{n} - a_i \overline{X})^2 = \sigma_u^2 \sum_{i=1}^{n} (\frac{1}{n} - a_i \overline{X})^2 = \sigma_u^2 \sum_{i=1}^{n} (\frac{1}{n^2} - 2\frac{a_i \overline{X}}{n} + a_i^2 (\overline{X})^2)$$

$$= \sigma_u^2 (\frac{n}{n^2} - 2\frac{\overline{X}\sum_{i=1}^{n} a_i}{n} + (\overline{X})^2 \sum_{i=1}^{n} a_i^2) = \sigma_u^2 (\frac{1}{n} + \frac{(\overline{X})^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2}) \text{ as } \sum_{i=1}^{n} a_i = 0 \text{ and } \sum_{i=1}^{n} a_i^2 = \frac{1}{\sum_{i=1}^{n} (X_i - \overline{X})^2}.$$

Now we have $\operatorname{var}(\hat{\beta}_2) = \frac{\sigma_u^2}{\sum (X_i - \overline{X})^2}$ and $\operatorname{var}(\hat{\beta}_1) = \sigma_u^2 \left\{ \frac{1}{n} + \frac{\overline{X}^2}{\sum \left(X_i - \overline{X}\right)^2} \right\}$. It remains to consider the covariance of the two estimators.

(e) Prove that $\operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) = \frac{-\overline{X}\sigma_u^2}{\sum (X_i - \overline{X})^2}$.

Solution

$$\operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) = \operatorname{cov}(\overline{Y} - \hat{\beta}_2 \overline{X}, \hat{\beta}_2) = \operatorname{cov}(\overline{Y}, \hat{\beta}_2) - \overline{X} \operatorname{cov}(\hat{\beta}_2, \hat{\beta}_2) = 0 - \overline{X} \operatorname{var}(\hat{\beta}_2) = \frac{-\overline{X}\sigma^2}{\sum (X_i - \overline{X})^2}.$$

The equality $\operatorname{cov}(\overline{Y}, \hat{\beta}_2) = 0$ can be proven in the following way

$$\operatorname{cov}(\overline{Y}, \hat{\beta}_2) = \operatorname{cov}\left(\sum \frac{1}{n} Y_i, \sum a_i Y_i \right) = \sum \left(\frac{a_i}{n} \right) \operatorname{var}(Y_i) + \sum_{i<j} \left(\frac{a_j}{n} \right) \operatorname{cov}(Y_i, Y_j) = \frac{\sigma_u^2}{n} \sum a_i = 0$$

as $\operatorname{cov}(Y_i, Y_j) = 0$ where $i \neq j$.

We use here the decomposition $\hat{\beta}_2 = \sum a_i Y_i$, where $a_i = \frac{(X_i - \overline{X})}{\sum (X_i - \overline{X})^2}$, and $\sum a_i = 0$.

Question 3. A

1. A researcher has data on the hourly earnings, E, measured in U.S. \$, tenure with current employer, T, in years, and age, AGE, also in years for 112 male and female high school graduates (with no further education) and 28 high school drop-outs. Defining a dummy variable D to be equal to 0 in the case of a graduate and 1 in the case of a drop-out, and a slope dummy DT as the product of D and T, the researcher runs the following regressions (standard errors in parentheses; RSS = sum of squares of residuals):

$$\textit{Graduates only} \quad \hat{E}_i = \underset{(1.70)}{14.82} + \underset{(0.14)}{0.70}T_i \qquad R^2 = 0.19 \quad RSS = 15{,}220 \qquad \text{(eq.1)}$$

$$\textit{Drop-outs only} \quad \hat{E}_i = \underset{(0.94)}{7.41} + \underset{(0.16)}{0.47}T_i \qquad R^2 = 0.24 \quad RSS = 294 \qquad \text{(eq.2)}$$

$$\textit{Combined sample} \quad \hat{E}_i = \underset{(1.39)}{12.26} + \underset{(0.12)}{0.81}T_i \qquad R^2 = 0.24 \quad RSS = 17{,}022 \qquad \text{(eq.3)}$$

$$\textit{Combined sample} \quad \hat{E}_i = \underset{(1.52)}{14.94} + \underset{(0.12)}{0.70}T_i - \underset{(2.34)}{8.46}D_i \qquad R^2 = 0.31 \quad RSS = 15{,}535 \qquad \text{(eq.4)}$$

$$\textit{Combined sample} \quad \hat{E}_i = \underset{(1.55)}{14.82} + \underset{(0.12)}{0.70}T_i - \underset{(3.35)}{7.41}D_i - \underset{(0.53)}{0.23}DT_i \qquad R^2 = 0.31 \quad RSS = 15{,}514 \qquad \text{(eq.5)}$$

(a) Give an economic interpretation of the last equation. Derive from the last equation two regression equations stating the relationship between T and E separately for graduates and for drop-outs.

a) The constant estimates the earnings of a high school graduate at the beginning of her/his tenure with a current employer as \$14.82 per hour
The coefficient of T estimates the increase in the hourly rate to be \$0.70 per year of the current tenure
The coefficient of D estimates that drop-outs earn \$7.41 less per hour, for any fixed level of T (or holding T constant)
The coefficient of the slope dummy estimates that the hourly earnings of drop-outs increase by \$0.23 less per year for drop-outs

Mathematically, the relationship between T and E corresponding to eq.5 is:

$$E_i = \beta_1 + \beta_2 T_i + \beta_3 D_i + \beta_4 DT_i + u_i$$

For graduates $D = 0$ so $\quad E_i = \beta_1 + \beta_2 T_i + u_i$

For drop-outs $D = 1$ so $\quad E_i = \beta_1 + \beta_2 T_i + \beta_3 1 + \beta_4 1 \cdot T_i + u_i =$

$$E_i = (\beta_1 + \beta_3) + (\beta_2 + \beta_4) T_i + u_i$$

According to this we get

Graduates $\quad \hat{y} = 14.82 + 0.70T$

Drop-outs $\quad \hat{y} = (14.82 - 7.41) + (0.70 - 0.23)T = 7.41 + 0.47T$

(b) Investigate whether the earnings function for drop-outs differs from that for graduates using a Chow test. State the null hypothesis; indicate the test used, degrees of freedom, and critical values. Explain the logic of the Chow test.

c) The null hypothesis that the parameters are the same for graduates and drop-outs.
Writing RSS_G, RSS_D and RSS_P for the residual sums of squares in the equations (eq.1), (eq.2) and (eq.3), the F statistic for the Chow test is

$$F(k+1, n-2k-2) = \frac{RSS_P - [RSS_G + RSS_D]/k}{(RSS_G + RSS_D)/(n-2k)}$$

where k is the number of parameters in the regression.
Since k is 2 and n is 140 in this case, the F statistic is

$$F(2,140) = \frac{(17{,}022 - [15{,}220 + 294])/2}{(15{,}220 + 294)/140 - 2 \cdot 2} = 6.61\,.$$

The critical value of F at the 1% significance level is about 4.79. Hence the null hypothesis is rejected and one concludes that the earnings functions for graduates and drop-outs are significantly different

(c) Investigate the same problem testing the explanatory power of the dummy variables. Perform the t-tests and the appropriate F-test (indicate what test should be performed). Compare the findings of the two test approaches. Discuss the results obtained from different approaches.
=0.23/0.53

c) t-tests: The *t* ratios for *D* and *DT* are -7.41/3.35 = -2.21 and 0.23/0.53 = -0.43, respectively. The first is significant at 5% level (critical value for df=120 is 1.98), while the second is not.
F test of the joint explanatory power of the dummy variables: Writing the coefficients of D and DT as γ_1 and γ_2, respectively, the null hypothesis is $H_0: \gamma_1 = \gamma_2 = 0$ and the alternative hypothesis is that one or both coefficients is non-zero. Writing the residual sums of squares in the third and fifth equations as RSS_3 and RSS_5, the F statistic is

$$F(2;1{,}064) = \frac{(RSS_3 - RSS_5)/2}{RSS_5/136} = \frac{(17{,}022 - 15{,}514)/2}{15{,}514/136} = 6.61\,.$$

This of course is the same F statistic as in the case of the Chow test, and so the conclusion is the same, the null hypothesis being rejected.
The Chow test and the F test on the dummy variables as a group are equivalent.
If the null hypothesis is rejected, the dummy variable approach may identify which coefficient is significantly different via t tests on the individual coefficients. In this case, only one t test is significant, possibly as a consequence of multicollinearity.

The researcher suspects that age, AGE, influence the hourly earnings, so she defines a new variable $A_i = AGE_i - \overline{AGE}$ where $\overline{AGE} = 38.15$ and includes it in the equation for the combined sample in the following way:

$$\hat{E}_i = \underset{(1.57)}{17.50} + \underset{(0.12)}{0.60}T_i - \underset{(2.34)}{5.22}D_i + \underset{(0.08)}{0.24}A_i - \underset{(0.005)}{0.02}A_i^2 \qquad R^2 = 0.39 \quad RSS = 13{,}707 \qquad \text{(eq.6)}$$

(d) How can be interpreted the intercept of this regression? Why not just A_i, but also its square A_i^2 are included in the equation? What is the meaning of their coefficients? Are the variables A_i and A_i^2 taken together significant?

d) Intercept equal to 17.5 shows the estimated hourly earnings of the high school graduates aged 38.15 at the beginning of their current tenure.

The presence of the square of demeaned (centered) age makes the dependence nonlinear: in addition to the general trend growth of hourly earnings with the growing age there is a penalty in the form of its square, which reduces the earnings of older workers and young people.

The demeaned age A and its square A^2 change simultaneously, so the marginal effect of one additional year can be measured using differential of a function $f(A) = 0.24A - 0.02A^2$ that is $df(A) = 0.24 - 2 \cdot 0.02A$. For example for the person aged 45($A = 45 - 38.15 = 6.85$) the marginal effect of additional year will be estimated as $dE = 0.24 - 0.04 \cdot 6.85 = -0.034$ so the positive trend, and the penalty for deviation of age from the mean (the age of most effective employees) are mutually canceled each other.

F test of the joint explanatory power of the age variables: Comparing RSS_3 and RSS_6 (for the eq.6), we get F statistic as

$$F(2;1{,}064) = \frac{(RSS_3 - RSS_6)/2}{RSS_6/136} = \frac{(17{,}022 - 13{,}707)/2}{13{,}707/136} = 16.45$$

what is certainly significant at 1% level.

To make her analysis more detailed she defines a group of dummy variables, $A20$ equal 1 only for people aged less than 30, and 0 for the others, $A30$ equal 1 only for people aged 30 to 40, and 0 for the others, $A40$ equal 1 only for people aged 40 to 50, and 0 for the others, and $A50$ equal 1 only for people aged from 50 and more, and 0 for the others. Then she runs the regression using combined sample

$$\underset{}{\hat{E}_i} = \underset{(2.12)}{9.82} + \underset{(0.11)}{0.55}T_i - \underset{(2.33)}{4.37}D_i + \underset{(2.44)}{4.81}A30_i + \underset{(2.37)}{10.97}A40_i - \underset{(3.11)}{1.26}A50_i \qquad R^2 = 0.43 \quad RSS = 12{,}878 \qquad \text{(eq.7)}$$

(e) How can be interpreted the intercept of this regression? What is the meaning of coefficients of dummy variables connected with age? Why the researcher did not include also the variable $A20$ into equation? Is any sense in changing reference category here running new regressions?

d) Intercept equal to 9.82 now shows the estimated hourly earnings of the high school graduates aged less than 30 at the beginning of their current tenure.

Coefficients of age variables show premium for the persons belonging to corresponding age category, for example, 10.97 shows that people of age from 40 to 50 get 10.97 more than young workers (aged less than 30) for any fixed length of tenure (this effect is significant). As we can see people aged more than 50 get even less than young people (by 1.26), but this effect is insignificant.

Inclusion in the equation at the same time all the dummies would lead to perfect multicollinearity, which would make the regression estimation impossible.

For groups of dummy with two categories the changing of reference category leads only to a change in the sign of the coefficient of the swapping variables, that does not give additional information.

In the case of more than two categories the absolute values and significance of the coefficients give information only in comparison with reference category. Changing reference category leads to changes in the values of coefficients and their significance, what gives new information. For example in (eq.7) there is significant difference in the earnings between people aged 40-50 and people aged less than 30, but we have no information from this equation whether the difference between people aged 40-50 and people aged 30-40 (or people aged more than 50) is significant or not. Taking people aged 40-50 as reference category we can get the answer to this question.

The International College of Economics and Finance
Econometrics. Mid-term exam. 2014 October 30

Part 2. (1 hour 30 minutes). Answer the first (obligatory) question and one of the questions 2 or 3.

IMPORTANT: Start answering each question from the new page (ask for extra paper if necessary). Structure your answers in accordance with the structure of the questions. Testing hypotheses always state clearly null and alternative hypotheses, provide critical value used for test, mentioning degrees of freedom and the significance level chosen for the test.

Answer all five parts of question 1

1. all parts of this question are obligatory.

a) A student of a local university prepared the small econometric project devoted to 'Decline of an old railway' based on the data of the expenditures of the citizens of a remote mountainous province on the local railway transportation *RAIL* (in thousands of dollars) and aggregate personal disposable income of the citizens of the province *DPI* (in hundreds of thousands of dollars), and has obtained the following equation (14 weekly observations):

$$RAIL = 395.37 - 0.078 DPI$$
$$(s.e.) \quad (24.16) \quad (0.036)$$

In his report he wrote: «at least one of the coefficients in my regression is significant, so the decline of mountain railways is inevitable: the growth of income of local population leads to the decrease of using traditional ways of transportations. According my regression annual decrease of expenditures on mountain railway counts to 0,078 percent'.

a1) Do you agree with student's analysis? If not, correct his statements?

Solution and marking

a1) The null-hypothesis for the testing significance of the equation of the simple linear regression $RAIL = \beta_1 + \beta_2 DPI + u$ is $H_0 : \beta_2 = 0$, the significance of intercept has nothing to do with the significance of the regression equation. In fact $H_0 : \beta_2 = 0$ cannot be rejected as

$t = \dfrac{b_2}{s.e.(b_2)} = \dfrac{-0.078}{0.036} = -2.176$, while $t(crit, 2sided, df = 14-2, 5\%) = 2.179$.

So we cannot tell here on the 'inevitable decline' of the expenditures on transportation with the increase of income. The correct interpretation (in the case if it were significant): with increase of income by one hundred of thousands of dollars the expenditures on railway transportation falls by 78 dollars. The interpretation of the student is nonsense.

[3 marks].

a2) Using method of confidence intervals show that the student's results are compatible with the hypothesis that the true value of the coefficient of *DPI* is positive.

Solution and marking

a1) The 5% confidence interval for β_2 can be constructed as

$$(b_2 - s.e.(b_2)\cdot t(cr., df = 12, 5\%);\ b_2 + s.e.(b_2)\cdot t(cr., df = 12, 5\%))$$

or

$$(-0.078 - 0.036\cdot 2.179;\quad -0.078 + 0.036\cdot 2.179)$$

or

$$(-0.1564;\quad 0.000444)$$

This confidence interval includes 0 and some positive numbers.
As we know the confidence interval is a set of all possible null-hypotheses compatible with the data. So the results are compatible with the hypothesis that the true value of the coefficient of *DPI* is positive. **[3 marks]**.

a3) The student forgot to include into equation the value of determination coefficient R^2. Try to restore it based on the regression results and comment on its value.

Solution and marking

From a) $t = \frac{b_2}{s.e.(b_2)} = \frac{0.078}{0.036} = 2.176$, so we can evaluate F-statistic $F = t^2 = 2.176^2 = 4.73$

Now we can find R^2 from the equation $F = \frac{R^2}{(1-R^2)} \cdot (14-2)$, solving for R^2 we get $R^2 = 0.28$.

So only 28% of the variance of dependent variable are explained by the variable *DPI* included into equation. One should recommend the student to include more explanatory variables, for example prices for transportation and so on to improve overall fit of the equation.

[2 marks].

[Total 8 marks for c)].

b) To estimate the regression model $Y_i = \beta_1 + \beta_2 X_i + u_i$ based on 4 observations $(X_i; Y_i)$ the following results were obtained: $\sum Y_i^2 = 30$, $\sum X_i^2 = 54$, $\sum X_i Y_i = 39$, $\sum Y_i = 10$, $\sum X_i = 12$, $\sum x_i^2 = \sum (X_i - \overline{X})^2 = 18$. $\sum e_i^2 = 0.5$ (e_i - are residuals of the regression).

b1) Find OLS estimates of regression coefficients and their standard errors.

b2) Test the hypothesis $H_0: \beta_2 = 0$, against the alternative $H_a: \beta_2 \neq 0$..

b3) Does the result of the test change if one-sided alternative hypothesis $H_a: \beta_2 > 0$ is used for the test?

Solution

b1) $b_2 = \hat{\beta}_2 = \frac{n\sum X_i Y_i - (\sum X_i)(\sum Y_i)}{n\sum X_i^2 - (\sum X_i)^2} = \frac{4 \cdot 39 - 12 \cdot 10}{4 \cdot 54 - (12)^2} = \frac{1}{2}$.

$b_1 = \hat{\beta}_1 = \overline{Y} - b_2 \overline{X} = \frac{1}{n}(\sum Y_i) - b_2(\frac{1}{n}\sum X_i) = \frac{10}{4} - \frac{1}{2} \cdot \frac{12}{4} = 1$.

So the regression is $\hat{Y}_i = 1 + \frac{1}{2} X_i$. **[2 marks]**

Let us evaluate the standard error of coefficient b_2

$$\text{Var}(b_2) = \frac{s_u^2}{\sum x_i^2} = \frac{\frac{1}{n-2}\sum e_i^2}{\sum (X_i - \overline{X})^2} = \frac{\frac{0.5}{4-2}}{18} = \frac{1/4}{18} = \frac{1}{72}$$, **[1 marks]**

so

$$s.e.(b_2) = \sqrt{\frac{1}{72}} = \frac{1}{6\sqrt{2}} = 0.117$$ **[1 marks]**

$$s.e.(b_1) = \sqrt{\frac{s_u^2}{n}\left(1 + \frac{\overline{X}^2}{\mathrm{Var}(X)}\right)} = \sqrt{\frac{s_u^2}{n}\left(1 + \frac{n\overline{X}^2}{\sum (X_i - \overline{X})^2}\right)} = \sqrt{\frac{1}{4 \cdot 4}\left(1 + \frac{4 \cdot (3)^2}{18}\right)} = \frac{1}{4}\sqrt{\left(1 + \frac{2 \cdot 9}{9}\right)} = 0.433$$

[1 marks]

b2)

Now evaluate *t*-statistics: $t = \frac{b_2 - \beta_2}{s.e.(b_2)} = \frac{0.5 - 0}{0.117} = 4.26$

For two-sided alternative $4.26 < 4.303 = t(two\ sided, crit., 5\%, df = 2)$ - not significant. **[2 marks]**

b3) One-sided alternative $H_a : \beta_2 > 0$ helps to state the significance

$4.26 > 2.92 = t(one\ sided, crit., 5\%, df = 2)$.**[1 mark],**

[Total 8 marks for b)].

c) In the standard linear model $Y_i = \beta_1 + \beta_2 X_i + u_i$ assume that the error term u_i takes the value 0 with probability 0.5 and the value 1 with probability 0.5.

c1) Explain how, using Monte Carlo techniques, you could assess the effect of this distribution of the error term on the distribution of the ordinary least squares estimate of β_2.

c2) Suppose linear regression model will satisfy all the assumptions of the Model A except the condition that the random term is normally distributed. Instead it has distribution as described above. Are the OLS estimators still BLUE? Is it possible to test their significance using conventional t-tests and F-test?

Solution and marking

c1) Create a series X_t and choose values of β_1 and β_2. Now generate the error series u_t such that it takes a value 0 with probability 0.5 and 1 with probability 0.5 (you could toss a coin). From the information create the series Y_t and regress Y_t on X_t to estimate b_1 and b_2.

Repeat the process many times with the same X_t, β_1 and β_2.

Create a histogram of b_1 and b_2. These will show the distributions of the OLS estimates of β_1 and β_2.

The distributions will vary depending on the length of the X_t series. . **[5 mark]**

c2) The assumption of normality of disturbance term is not used in the proof of Gauss-Markov theorem, however E(u) not equals to zero anymore, as in binomial distribution E(u)=p*n, so it violates the condition of Gauss-Markov and therefore estimates are NOT BLUE anymore. **[2 mark]**
The distributions of statistics of t-test and F-test are essentially based on the normal distribution so now these tests are not more valid. **[1 mark]**.
[Total 8 marks for c)].

d) Let true model be $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$ and fitted model estimated by OLS be $\hat{Y} = b_1 + b_2 X_2 + b_3 X_3$.

d1) What is OLS estimator b_3 of multiple regression coefficient β_3? Is it really different from the estimator of coefficient β_3 in simple linear regression $Y = \beta_1 + \beta_3 X_3 + u$? Are the properties of these estimators also different? If yes what are the differences?

Solution

d1) The formula for estimator b_3 of the coefficient β_3 can be derived from OLS equations:

$b_3 = \dfrac{\text{Cov}(X_3,Y)\text{Var}(X_2) - \text{Cov}(X_2,Y)\text{Cov}(X_3,X_2)}{\text{Var}(X_3)\text{Var}(X_2) - [\text{Cov}(X_3,X_2)]^2}$, so the expression is different from OLS estimator $b_3^* = \dfrac{\text{Cov}(X_3,Y)}{\text{Var}(X_3)}$ for the simple linear regression $Y = \beta_1 + \beta_3 X_3 + u$. The meaning of the coefficient b_3 is also different from the meaning of coefficient b_3^*: in simple linear regression $\hat{Y} = b_1^* + b_3^* X_3$ b_3^* shows the marginal effect of X_3, while in multiple regression $\hat{Y} = b_1 + b_2 X_2 + b_3 X_3$ b_3 shows marginal effect of X_3 while X_2 being fixed. Variance of the coefficient b_3^* is given by expression

$\sigma_{b_3^*}^2 = \dfrac{\sigma_u^2}{n\text{Var}(X_3)}$ while variance of the coefficient b_3 is given by another expression

$\sigma_{b_3}^2 = \dfrac{\sigma_u^2}{n\text{Var}(X_3)} \cdot \dfrac{1}{1 - r_{X_2,X_3}^2}$, which is always not less than for the simple regression model.

[4 marks]

d2) Prove that OLS estimator b_3 of the coefficient β_3 in multiple regression is unbiased.

Solution

d2) As it is known

$$Eb_3 = \beta_3 + E\left(\frac{\text{Cov}(X_3,u)\text{Var}(X_2) - \text{Cov}(X_2,u)\text{Cov}(X_3,X_2)}{\Delta}\right)^{*)}$$

$$E(b_3) = \beta_3 + \frac{\text{Var}(X_2)E(\text{Cov}(X_3,u)) - \text{Cov}(X_3,X_2)E(\text{Cov}(X_2,u))}{\Delta} =$$

$$= \beta_3 + \frac{\text{Var}(X_2)(\text{Cov}(X_3,Eu)) - \text{Cov}(X_3,X_2)(\text{Cov}(X_2,Eu))}{\Delta} = \beta_3 + 0 = \beta_3$$

*) It can be proved like this

$$b_3 = \frac{1}{\Delta}(\text{Cov}(X_3,Y)\text{Var}(X_2) - \text{Cov}(X_2,Y)\text{Cov}(X_3,X_2)) =$$

$$= \frac{1}{\Delta}((\text{Cov}(X_3,\beta_1 + \beta_2 X_2 + \beta_3 X_3 + u)\text{Var}(X_2) - \text{Cov}(X_2,\beta_1 + \beta_2 X_2 + \beta_3 X_3 + u)\text{Cov}(X_3,X_2)) =$$

$$= \frac{1}{\Delta}((\beta_2\text{Cov}(X_3,X_2) + \beta_3\text{Var}(X_3) + \text{Cov}(X_3,u))\text{Var}(X_2) -$$
$$- \beta_2\text{Var}(X_2) - \beta_3\text{Cov}(X_3,X_2) - \text{Cov}(X_2,u))\text{Cov}(X_3,X_2)) =$$

$$= \frac{1}{\Delta}(\beta_2\text{Cov}(X_3,X_2)\text{Var}(X_2) + \beta_3\text{Var}(X_3)\text{Var}(X_2) + \text{Cov}(X_3,u)\text{Var}(X_2) -$$
$$- \beta_2\text{Var}(X_2)\text{Cov}(X_3,X_2) - \beta_3\text{Cov}(X_3,X_2)\text{Cov}(X_3,X_2) - \text{Cov}(X_2,u)\text{Cov}(X_3,X_2)) =$$

$$= \frac{1}{\Delta}(\beta_3\{\text{Var}(X_3)\text{Var}(X_2) - [\text{Cov}(X_3,X_2)]^2\} + \beta_2\{\text{Cov}(X_3,X_2)\text{Var}(X_2) - \text{Var}(X_2)\text{Cov}(X_3,X_2)\} +$$
$$+ \{\text{Cov}(X_3,u)\text{Var}(X_2) - \text{Cov}(X_2,u)\text{Cov}(X_3,X_2)\}) =$$

$$= \beta_3 \frac{\Delta}{\Delta} + \beta_2 \frac{1}{\Delta}\{\mathrm{Cov}(X_3, X_2)\mathrm{Var}(X_2) - \mathrm{Var}(X_2)\mathrm{Cov}(X_3, X_2)\} +$$

$$+ \frac{1}{\Delta}\{\mathrm{Cov}(X_3, u)\mathrm{Var}(X_2) - \mathrm{Cov}(X_2, u)\mathrm{Cov}(X_3, X_2)\}) =$$

So

$$E(b_3) = \beta_3 + 0 + \frac{1}{\Delta} E\{\mathrm{Cov}(X_3, u)\mathrm{Var}(X_2) - \mathrm{Cov}(X_2, u)\mathrm{Cov}(X_3, X_2)\}$$

[4 marks]

(e) Show that the OLS estimators of β_1 and β_2 for the following regression models are identical.

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad (1)$$

$$Y_t + X_t = \beta_1 + (\beta_2 + 1) X_t + u_t . \qquad (2)$$

Nevertheless, the R-squares for these regressions are, in general, different.

Solution:

To estimate regression (1) using OLS we have to solve the optimization problem $\sum (Y_t - b_1 - b_2 X_t)^2 \to \min$ for the regression $\hat{Y}_t = b_1 + b_2 X_t$. The corresponding minimization problem for the regression (2) $(Y_t \hat{+} X_t) = b_1 + (b_2 + 1) X_t$ is $\sum (Y_t + X_t - b_1 - b_2 X_t - X_t)^2 \to \min$, but it is identical to the problem $\sum (Y_t - b_1 - b_2 X_t)^2 \to \min$, so the solutions of both problems are the same.

So the estimators are the same.

It could be said additionally that the properties of regression coefficients estimators are fully determined by their expressions through the data and the expressions for the residuals (for example, the standard errors and test statistics based on them). But as it is known the residuals of the regression are described by the difference between the left and right side of the first OLS equation, so they are also the same. So two regressions (1) and (2) are equivalent: they give the identical estimators with the same properties.

As residuals are the same RSS's are also the same. But TSS's are different. It is sufficient to give an example: let $Y_1 = 3, Y_2 = 4, Y_3 = 5$, $X_1 = 1, X_2 = 2, X_3 = 3$, so $\overline{Y} = 4, \overline{X} = 2$ and $TSS_1 = \sum (Y_t - \overline{Y})^2 = 2$ while $TSS_2 = \sum (Y_t + X_t - \overline{Y} - \overline{X})^2 = 8$, so the determination coeffients for two regressions $R^2 = 1 - \frac{RSS}{TSS}$ should be also different. **[8 marks]**

Question 2.

2. A simple linear regression $Y_i = \beta_1 + \beta_2 X_i + u_i$ is considered where the regression equation and disturbance term u_i satisfy all assumptions of model A (so $\text{cov}(u_i, u_j) = 0, i \neq j$, and $\text{var}(u_i) = \sigma_u^2$ for $n = 1, 2, ..., n$). Let $\hat{\beta}_1$ and $\hat{\beta}_2$ are OLS estimators of β_1 and β_2 correspondingly.

a) Prove that the variance of OLS estimator of the slope coefficient is $\text{var}(\hat{\beta}_2) = \dfrac{\sigma_u^2}{\sum\left(X_i - \overline{X}\right)^2}$

Solution:

$$\sigma_{b_2}^2 = E\left\{(b_2 - E(b_2))^2\right\} = E\left\{(b_2 - \beta_2)^2\right\} = E\left\{\left(\sum_{i=1}^{n} a_i u_i\right)^2\right\} =$$

$$= E\left\{\sum_{i=1}^{n} a_i^2 u_i^2 + \sum_{i=1}^{n}\sum_{j \neq i} a_i a_j u_i u_j\right\} = \sum_{i=1}^{n} a_i^2 E\left(u_i^2\right) + \sum_{i=1}^{n}\sum_{j \neq i} a_i a_j E\left(u_i u_j\right) =$$

$$= \sum_{i=1}^{n} a_i^2 \sigma_u^2 = \sigma_u^2 \sum_{i=1}^{n} a_i^2 = \frac{\sigma_u^2}{\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2}$$

[7 marks]

b) Prove that $\text{cov}(\overline{Y}, \hat{\beta}_2) = 0$.

Solution

As it is known $\hat{\beta}_2 = \sum a_i Y_i$, where $a_i = \dfrac{(X_i - \overline{X})}{\sum (X_i - \overline{X})^2}$, and $\sum a_i = 0$. So

$$\text{cov}(\overline{Y}, \hat{\beta}_2) = \text{cov}\left(\sum \frac{1}{n} Y_i, \sum a_i Y_i\right) = \sum \left(\frac{a_i}{n}\right) \text{var}(Y_i) + \sum_{i \neq j} \left(\frac{a_j}{n}\right) \text{cov}(Y_i, Y_j) = \frac{\sigma_u^2}{n} \sum a_i = 0$$

as $\text{cov}(Y_i, Y_j) = 0$ where $i \neq j$.

[7 marks]

c) Prove that $\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = \dfrac{-\overline{X}\sigma_u^2}{\sum (X_i - \overline{X})^2}$.

Solution

$$\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = \text{cov}(\overline{Y} - \hat{\beta}_2 \overline{X}, \hat{\beta}_2) = \text{cov}(\overline{Y}, \hat{\beta}_2) - \overline{X}\,\text{cov}(\hat{\beta}_2, \hat{\beta}_2) = 0 - \overline{X}\,\text{var}(\hat{\beta}_2) = \frac{-\overline{X}\sigma^2}{\sum (X_i - \overline{X})^2}.$$

[7 marks]

d) Show that the variance of the predicted value of the regression $\text{var}(\hat{\beta}_1 + \hat{\beta}_2 X^*)$ is equal to $\left(\dfrac{1}{n} + \dfrac{(X^* - \overline{X})^2}{\sum\left(X_i - \overline{X}\right)^2}\right)\sigma_u^2$.

Solution

$$\operatorname{var}(\hat{\beta}_1+\hat{\beta}_2 X^*) = \operatorname{var}(\hat{\beta}_1)+(X^*)^2 \operatorname{var}(\hat{\beta}_2)+2X^* \operatorname{cov}(\hat{\beta}_1,\hat{\beta}_2).$$

As it is known

$$\operatorname{var}(\hat{\beta}_1)=\sigma_u^2\left\{\frac{1}{n}+\frac{\overline{X}^2}{\sum\left(X_i-\overline{X}\right)^2}\right\} \quad \operatorname{var}(\hat{\beta}_2)=\frac{\sigma_u^2}{\sum\left(X_i-\overline{X}\right)^2} \quad \operatorname{cov}(\hat{\beta}_1,\hat{\beta}_2)=\frac{-\overline{X}\sigma^2}{\sum(X_i-\overline{X})^2}. \text{ So}$$

$$\operatorname{var}(\hat{\beta}_1+\hat{\beta}_2 X^*)=\sigma_u^2\left(\frac{1}{n}+\frac{\overline{X}^2}{\sum\left(X_i-\overline{X}\right)^2}\right)+(X^*)^2\frac{\sigma_u^2}{\sum\left(X_i-\overline{X}\right)^2}+2X^*\frac{-\overline{X}\sigma^2}{\sum(X_i-\overline{X})^2}=$$

$$=\left(\frac{1}{n}+\frac{(\overline{X}^2-2X^*\overline{X}+(X^*)^2)}{\sum\left(X_i-\overline{X}\right)^2}\right)\sigma_u^2=\left(\frac{1}{n}+\frac{(X^*-\overline{X})^2}{\sum\left(X_i-\overline{X}\right)^2}\right)\sigma_u^2.$$

[7 marks]

By definition the prediction error for the regression model is $PE=Y^*-\hat{Y}^*$ where Y^* is the value of Y corresponding to the specified value X^* of explanatory variable X.

e) Show that the variance of the prediction error $\operatorname{var}(PE)$ is equal to $\left(1+\frac{1}{n}+\frac{(X^*-\overline{X})^2}{\sum(X_i-\overline{X})^2}\right)\sigma_u^2$.

Solution

$$\operatorname{var}(PE)=\operatorname{var}(Y^*-\hat{Y}^*)=\operatorname{var}(Y^*)+\operatorname{var}(\hat{Y}^*)-2\operatorname{cov}(Y^*,\hat{Y}^*).$$

Because we are predicting a future value Y^* that is not employed in the computation of $\hat{Y}^*$, it follows that Y^* and $\hat{Y}^*$ are independent and hence that $\operatorname{cov}(Y^*,\hat{Y}^*)=0$.

(This can be expressed also in more strict manner saying that by assumption the disturbance term of future value Y^* should be independent of disturbance terms $u_i, i=1, 2, ..., n.$ For more details see *V.P.Nosko – Econometrics for beginners. – 2.12*)

Then $\operatorname{var}(PE)=\operatorname{var}(Y^*)+\operatorname{var}(\hat{Y}^*)=\sigma^2+\operatorname{var}(\hat{\beta}_1+\hat{\beta}_2 X^*)=$

$$=\sigma_u^2+\left(\frac{1}{n}+\frac{(X^*-\overline{X})^2}{\sum(X_i-\overline{X})^2}\right)\sigma_u^2=\left(1+\frac{1}{n}+\frac{(X^*-\overline{X})^2}{\sum(X_i-\overline{X})^2}\right)\sigma_u^2.$$

[7 marks]

3. A researcher has quarterly data on the expenditure on jewelry EOJ_t for 6 consecutive years (total 24 observations) in some developed country (in billions of euro). He considers disposable personal income DPI_t (also in billions of euro) and index of relative prices on jewelry PRJ_t as determinants of household spending on this category of goods and first runs regression (standard errors are in parentheses here and further in all equations)

$$\log(\hat{EOJ})_t = \underset{(0.48)}{2.06} + \underset{(0.0002)}{0.001} \cdot DPI_t - \underset{(0.0033)}{0.015} \cdot PRJ_t \qquad R^2 = 0.9806, \quad RSS = 0.0738 \qquad \text{(eq.1)}$$

a) Give interpretation to the coefficients of equation and prove mathematically that your interpretation is correct. Write theoretical equation corresponding to the estimated equation showing explicitly the dependence of variable EOJ_t on the explanatory variables and disturbance term (OLS is used for estimation of this and all others equations).

Solution

For simplicity denote EOJ by Y, DPI by X_2, and PRJ by X_3. Then semilog regression fuction can be written as $\hat{\log Y} = b_1 + b_2 X_2 + b_3 X_3$.

To receive a verbal interpretation for coefficient b we must take the partial differential of the function $d(\log Y) = d(b_1 + b_2 X_2 + b_3 X_3) = b_2 dX_2$ so $\frac{dY}{Y} = b_2 dX_2$. Expressing b_2 from here we get $b_2 = \frac{\frac{dY}{Y}}{dX_2}$ or as relative gain is usually expressed as a percentage $b_2 \cdot 100\% = \frac{\frac{dY}{Y} \cdot 100\%}{dX_2}$.

Setting increment of X_2 to the unit we obtain the interpretation: if one multiply b_2 by 100, the resulting number shows the percentage increase of Y when X_2 will increase by one unit. In our case when *DPI* increases ny one billion of euro

To get explicit dependence of variable EOJ_t on the explanatory variables and disturbance term we first write the theoretical equation $\log(EOJ_t) = \beta_1 + \beta_2 \cdot DPI_t + \beta_3 \cdot PRJ_t + u_t$.

Potentiating obtain $EOJ_t = e^{\beta_1 + \beta_2 \cdot DPI_t + \beta_3 \cdot PRJ_t + u_t} = e^{\beta_1} e^{\beta_2 \cdot DPI_t} e^{\beta_3 \cdot PRJ_t} e^{u_t}$. The disturnabce term is now multiplicative and hace lognormal distribution (its logarithm has normal distribution).

[7 marks]

b) The researcher decided to take into account the index of relative prices on housing PRH_t as housing along with jewelry can be considered as an alternative way of investments, and got equation

$$\log(\hat{EOJ})_t = \underset{(0.95)}{2.91} + \underset{(0.0002)}{0.001} \cdot DPI_t - \underset{(0.0034)}{0.014} \cdot PRJ_t - \underset{(0.0082)}{0.0087} \cdot PRH_t \qquad R^2 = 0.9816, \quad RSS = 0.0670 \qquad \text{(eq.2)}$$

The researcher asked you to help him to evaluate the contribution of the variable PRH_t into equation using adjusted R-squared and some other tests. Perform calculations and comment on the results.

Solution

R^2 is the proportion of variance explained by the model, i.e. ESS/TSS where ESS is explained sum of squares and TSS is total sum of squares.
R^2 gives a measure of the overall "fit" of the model but it is not perfect: if more variables are included in the regression the value of R^2 will increase.
$\overline{R}^2$ is defined as $R^2 - \frac{k-1}{n-k}(1-R^2)$ or $1-\frac{n-1}{n-k}(1-R^2)$ where n is the sample size and k is the number of parameters.
Unlike R^2 the adjusted determination coefficient $\overline{R}^2$ has a penalty in the form of $-\frac{k-1}{n-k}(1-R^2)$ which will reduce the value as k increases.
For the equation (1) $\overline{R}^2 = 0.9806 - \frac{2}{(24-3)} \cdot (1-0.9806) = 0.9788$
For the equation (2) $\overline{R}^2 = 0.9816 - \frac{3}{(24-4)} \cdot (1-0.9816) = 0.9789$
The increase is very small or even dubious (the sign of increase depends on the accuracy of rounding).
$\overline{R}^2$ will increase if the absolute value of the t statistic of the included variable is greater than 1, i.e. it does not mean that the one or more coefficients become significant and so it does not mean that specification of the model has improved.
In fact the included variable *PRH* has insignificant coefficient so it is not reasonable include it into equation.
F-test for the contribution of one variable *PRH* is equivalent to t-test above so it cannot give additional information.
[6 marks]

It is known that in the middle of the fourth year (of the period under consideration) there was a banking crisis in the country that could affect trade in jewelry. So the researcher defines variable CR_t equal 0 in the pre-crisis period (14 quarterly observations) and 1 from the beginning of the crisis to the end of the period (10 quarterly observations). Then he includes the variable CR_t into original equation (1).

$$\log(\hat{EOJ})_t = \underset{(0.54)}{2.03} + \underset{(0.0002)}{0.001} \cdot DPI_t - \underset{(0.0032)}{0.015} \cdot PRJ_t - \underset{(0.048)}{0.061} \cdot CR_t \qquad R^2 = 0.982, \quad RSS = 0.0683 \qquad \text{(eq.3)}$$

c) How interpretation of the coefficients has changed with the inclusion of variable CR_t? Explain, why the variable CR_t could be insignificant, despite the obvious presence of the crisis for the researcher.

Solution
Writing the theoretical equation corresponding equation (3)

$$\log(EOJ_t) = \beta_1 + \beta_2 \cdot DPI_t + \beta_3 \cdot PRJ_t + \beta_4 \cdot CR_t + u_t$$

And setting $CR_t = 0$ (pre-crisis period) we get

$$\log(EOJ_t) = \beta_1 + \beta_2 \cdot DPI_t + \beta_3 \cdot PRJ_t + u_t$$

So now intercept $\beta_1 + \beta_2 \cdot DPI_t + \beta_3 \cdot PRJ_t$ is in fact intercept for pre-crisis period while β_4 shows the increase (decrease in our case) for the certain values of *DPI* and *PRJ* in post-crisis period as to compare with the same values of *DPI* and *PRJ* in pre-crisis period. The coefficients β_2 and β_3 has the same meaning as it was explained in (a) under implicit assumption that they are have same values in pre-crisis and post-crisis period.
Summarizing we have

$$\widehat{\log(EOJ)}_t = 2.03 + 0.001 \cdot DPI_t - 0.015 \cdot PRJ_t \quad \text{(pre-crisis period)}$$

$$\widehat{\log(EOJ)}_t = 1.969 + 0.001 \cdot DPI_t - 0.015 \cdot PRJ_t \quad \text{(post-crisis period)}$$

The coefficient of variable CR_t is insignificant ($t = (-0.06)/0.048 = -1.27$). Of course it can tell of the absence of the crisis, but more plausible that the assumption of equal values of the slope coefficients for pre-crisis and post-crisis periods was not true.

[6 marks]

Then he defines the full set of slope dummies, adds them into equation and runs new regression with the full set of dummies: (eq.4)

$$\widehat{\log(EOJ)}_t = 1.69 - 0.0016DPI_t - 0.0138PRJ_t - 0.05CR_t - 0.0006DPI*CR_t + 0.0057PRJ*CR_t \quad R^2 = 0.9890$$

$$(0.84)\ (0.00039) \qquad (0.005) \qquad (0.964) \qquad (0.0004) \qquad (0.0066) \qquad RSS = 0.04196$$

(eq.4).

d) Does this equation indicate on the presence of crisis? Is there any contradiction between the results of equations (3) and (4).

The colleague of the researcher noticed that it would be better to analyze data using sub-samples on pre-crises and post-crises sub-periods and then to perform Chow test. He sent him a paper where he obtained the regressions $\widehat{\log(EOJ)}_t = b_1 + b_2 DPI_t + b_3 PRJ_t$ with RSS_1 for the pre-crisis period equal to 0.03497 and RSS_2 for the post-crisis period equal to 0.006995. Explain what is Chow test, and perform calculations.

<u>Solution</u>

For the evaluation of the contribution of three crisis dummy variables we can use F-test comparing RSS_U for the unrestricted (*U*) equation (4) with RSS_R for the restricted (*R*) original equation (4) using statistic

$$F = \frac{(RSS_R - RSS_U)/3}{RSS_U/(24-6)}$$

This statistic has F-distribution with (3, 18) degrees of freedom. Let us evaluate it

$$F = \frac{(0.0738 - 0.04196)/3}{0.04196/(24-6)} = 4.55$$

$F(crit, df1 = 3, df2 = 18, 5\%) = 3.16$, $F(crit, df1 = 3, df2 = 18, 1\%) = 5.09$

So the group of crisis variables is significant only at 5% significance level.

Chow test compares RSS_p (pooled sample) (the same as RSS_R) for the equation (1) with the sums of squared residuals for two subsamples RSS_1, and RSS_2, corresponding F-statistic is

$$F = \frac{(RSS_p - (RSS_1 + RSS_2))/3}{(RSS_1 + RSS_2)/(24 - 2\cdot 3)}$$

Let us evaluate it

$$F = \frac{(0.0738 - (0.03497 + 0.006995)/3}{(0.03497 + 0.006995)/(24-6)} = 4.55$$

The critical values are the same

$F(crit, df1 = 3, df2 = 18, 5\%) = 3.16$, $F(crit, df1 = 3, df2 = 18, 1\%) = 5.09$

So the difference between analysis on the base of pooled sample and the detailed analysis based on subsamples is significant only at 5% significance level.
[6 marks]

e) Prove mathematically that Chow test is equivalent to the F-test for the full set of intercept and slope dummies (it is sufficient here to do this for the case under consideration).

Solution

We will use simple notations like in (a):

$$Y = \beta_1 + \beta_2 \cdot X_{2t} + \beta_3 \cdot X_{3t} + u_t \qquad \text{(e1)}$$

for the original equation with 2 explanatory variables, and

$$Y = \beta_1 + \beta_2 \cdot X_{2t} + \beta_3 \cdot X_{3t} + \beta_4 \cdot CR_t + \beta_5 \cdot X_{2t} \cdot CR_t + \beta_6 \cdot X_{3t} \cdot CR_t + u_t \qquad \text{(e2)}$$

for equation with full set of dummies. Note that for the first (pre-crisis) period equation (e2) under $CR_t = 0$ collapses to equation (e1) estimated for the sample of observations 1-14:

$$Y = \beta_1 + \beta_2 \cdot X_{2t} + \beta_3 \cdot X_{3t} + u_t \qquad \text{(t=1, 2, …,14)} \qquad \text{(e1-1)}$$

For the second (post-crisis) part of period equation (e2) under $CR_t = 1$ converts into equation

$$Y = (\beta_1 + \beta_4) + (\beta_2 + \beta_5) \cdot X_{2t} + (\beta_3 + \beta_6) \cdot X_{3t} + u_t \quad .$$

that is again into equation of the type (e1):

$$Y = \beta_1 + \beta_2 \cdot X_{2t} + \beta_3 \cdot X_{3t} + u_t \qquad \text{(t=15, 16,…,24)} \qquad \text{(e1-2)}$$

F-test for the group of all dummy (crisis) variables compares RSS_p (previous RSS_R) (sum of squared residuals for the estimated equation (e1) for the pooled sample of both periods) with RSS_d (previous RSS_U) (sum uf squared residuals for the estimated equation (e2) with full set of dummies for the same sample), corresponding F-statistic is

$$F = \frac{(RSS_p - RSS_d)/3}{RSS_d/(24-6)} \qquad \text{(e3)}$$

This statistic has F-distribution with (3, 18) degrees of freedom.
Chow test compares RSS_p with the sums of squared residuals of equation (e1-1) estimated for the first (pre-crisis) period - RSS_1, and of equation (e1-2) estimated for the second (post-crisis) period - RSS_2, corresponding F-statistic is

$$F = \frac{(RSS_p - (RSS_1 + RSS_2))/3}{(RSS_1 + RSS_2)/(24 - 2 \cdot 3)} \qquad \text{(e4)}$$

This statistic also has F-distribution with (3, 18) degrees of freedom.
So for euqivalence of both tests it is sufficient to prove that $RSS_d = RSS_1 + RSS_2$.
But it is obvious: as it was noted above for the pre-crisis period equation (e2) coincides with equation (e1-1) estimated for the data (1-14), so their residuals are identical, this means that for observations 1-14 first part of the sum of squared residuals of RSS_d is equal to RSS_1. For the post-crisis period equation (e2) coincides with equation (e1-2) estimated for the data (15-24), so their residuals are identical, this means that for observations 15-24 second part of the sum of squared residuals of RSS_d is equal to RSS_2. So $RSS_d = RSS_1 + RSS_2$, and both F-statistics are equal and have identical istribution. So the tests are equivalent (always give identical results).

[10 marks]

The International College of Economics and Finance
Econometrics – 2015-2016.
Mid-year exam. October 29.
Solutions.

IMPORTANT: Start answering each question from the new page (ask for extra paper if necessary). Structure your answers in accordance with the structure of the questions. Testing hypotheses always state clearly null and alternative hypotheses, provide critical value used for test, mentioning degrees of freedom and the significance level chosen for the test.

SECTION A

Answer **ALL** questions 1-3 from this section. (All questions from this section bear equal marks)

1. Working on her coursework a student of ICEF collected data on a sample of 247 ICEF graduates from different years of graduation working in Russia. She is interested in studying their current average monthly earnings – ***EARN*** (in thousands of rubles per month). Explanatory variables are ***G*** (average grade - from 0 to 10 - at the moment of graduation), ***EXP*** (work experience in years after graduation of a respondent), ***EXP2*** (work experience squared: $EXP2 = EXP^2$), and also some dummy variables: ***MDA*** (Master Degree Abroad – it is equal to 1 for those graduates who have received a master's degree abroad, and 0 otherwise), ***NFE*** (no further education - equal 1 for those graduates who received master's degree neither abroad nor in the country), and ***MALE*** (equal 1 for male and 0 for female). Here you have the results of estimation of 2 regressions using different sets of variables (standard errors in brackets).

$$\underset{}{\hat{EARN}_i} = \underset{(0.18)}{15.18} + \underset{(0.03)}{0.12}G_i + \underset{(0.021)}{0.254}EXP_i - \underset{(0.00005)}{0.001}EXP2_i \qquad R^2 = 0.68 \qquad (1)$$

$$\hat{EARN}_i = \underset{(0.049)}{10.068} + \underset{(0.04)}{0.1}G_i + \underset{(0.03)}{0.18}EXP_i - \underset{(0.00003)}{0.0008}EXP2_i + \underset{(0.456)}{4.021}MDA_i - \underset{(2.237)}{5.123}NFE_i + \underset{(0.016)}{0.025}MALE_i \qquad R^2 = 0.73 \qquad (2)$$

(a) How many categories of education level of ICEF graduates describe dummy variables ***MDA*** and ***NFE***? What is the reference category for two groups of dummies in the equation (2)? Why coefficients of common variables are different in equations (1) and (2)? What is the reason to include variable ***EXP2*** into equations?

(a) Three categories: no further education, master's degree in Russia and master's degree abroad.
The reference category corresponds to the values of all dummy variables equal to zero, so it is female person who got master's degree in Russia.
Equation (2) allows to describe the earnings functions for different categories of ICEF graduates (males and females, with master's degree abroad or not etc., while equation (1) describes earnings function of all ICEF graduates.
Coefficients of ***G***, ***EXP*** and ***EXP2*** represent partial marginal effect of abilities and work experience (that are the same for all categories defined by dummies). Inclusion of additional variables changes coefficients of regression. Inclusion of ***EXP2*** into equation allows to take into account the decreasing marginal productivity of ***EXP.***

(b) What is the difference between interpretations of equations (1) and (2) (give formal interpretation not paying attention to the significance of coefficients). Was it reasonable to include dummy variables based on the comparison of adjusted R-squared? Are the coefficients of the variables ***MDA***, ***NFE*** and ***MALE*** significant? Are they jointly significant?

(b) In equation (1) the constant term 15.18 gives initial year earnings for the ICEF graduate with zero experience (at the moment of graduation) and (with reservation) the lowest possible rating (at the moment of graduation). The intercept of the second equation 10.068 relates to the starting earnings of a 'fresh' **female** ICEF graduate with minimal rating ('fresh' means with no work experience) **who got master's degree in Russia**. Annual increase in earnings of a graduate with experience EXP_i according equation (1) is $\frac{\partial EARN}{\partial EXP} = 0.254 - 2 \cdot 0.001 EXP_i = 0.254 - 0.002 EXP_i$, while according equation (2) the annual increase is

$0.18 - 0.0016EXP_i$. Two estimates of marginal effects of ***EXP*** are different as equation (1) lacks many important variables. The interpretation of ***G*** coefficient is standard as marginal effects of additional final grade on earnings keeping other variables constant. The same for dummies. For example coefficient 4.021 of ***MDA*** means that those graduates who obtain their master degrees abroad get premium 4.021 thousand of rubles monthly as to compare to the graduates who obtain their master degrees in Russia, keeping values of other variables (***G, EXP***, ***MALE*** and of course ***NFE*** constant)..

Comparison of adjusted $\overline{R}^2 = R^2 - (1-R^2)\dfrac{k-1}{n-k}$:

$$\overline{R}_r^2 = 0.68 - (1-0.68)\frac{3}{247-4} = 0.676,$$

$$\overline{R}_u^2 = 0.73 - (1-0.73)\frac{7}{247-7} = 0.722$$

Some modest increase indicates that inclusion of three dummy variables is reasonable.

Absolute values of t-statistics for these variables are 8.82, 2.29, 1.56 while $t_{crit}^{1\%}(240) \approx t_{crit}^{1\%}(200) = 2.601$, $t_{crit}^{5\%}(240) \approx t_{crit}^{5\%}(200) = 1.972$ so the coefficient of ***MDA*** is significant at 1%, the one of ***NFE*** is significant only at 5% while the coefficient of ***MALE*** is insignificant.

Joint significance of the group of dummies. Finding F-statistics $F = \dfrac{(R_u^2 - R_r^2)/3}{(1-R_u^2)/240} = \dfrac{(0.73-0.68)/3}{(1-0.73)/240} = 14.815$ and comparing it with $F_{crit}^{1\%}(3, 240) \approx F_{crit}^{1\%}(3, 200) = 3.88$ we conclude that these variables as a group are significant.

(c) Is there any difference between earnings functions for men and women and how would you develop the model to investigate it?

(c) Equation (2) allows to get two different earnings function for men

$$\widehat{EARN}_i = (10.068 + 0.025) + 0.1G_i + 0.18EXP_i - 0.0008EXP2_i + 4.021MDA_i - 5.123NFE_i$$

and women

$$\widehat{EARN}_i = 10.068 + 0.1G_i + 0.18EXP_i - 0.0008EXP2_i + 4.021MDA_i - 5.123NFE_i$$

with a premium 0.025 for men, but difference is insignificant as coefficient of $MALE_i$ in **(b)** turned to be insignificant. It does not mean that there is no discrimination of women in earnings: the structure (or specification) of earnings function is underdeveloped. To develop it further one should introduce slope dummies $MALE_i \cdot G_i$, $MALE_i \cdot EXP_i$, $MALE_i \cdot EXP2_i$ and include them into equation on a par with intercept dummy $MALE_i$. Strictly speaking interactive dummies are also of use: $MALE_i \cdot MDA_i$ and $MALE_i \cdot NFE_i$. So the developed earnings function is

$$\widehat{EARN}_i = \beta_1 + \beta_2 MALE_i + (slope\ dummies) + (interactive\ dummies) + (control\ variables)$$

Now the earning function for men is

$$\widehat{EARN}_i = (\beta_1 + \beta_2) + (slope\ dummies) + (interactive\ dummies) + (control\ variables) \quad RSS1$$

while for women it has restricted form

$$\widehat{EARN}_i = \beta_1 + (control\ variables) \quad RSS2.$$

To evaluate the distinction (or rather discrimination) between men and women we have to evaluate the test (an ***F*** test based on comparison ***RSS*1** and ***RSS*2**) for the joint significance of all dummies connected with gender. This function will be much more sensitive so there is a chance that it will be significant.
Of course this approach is equivalent to Chow test for subsamples of men and women.

2. The researcher investigates the relation between GDP – variable Y_t and money supply (M1) – variable M_t for Canada in millions of dollars for a period 1989-2013 (in the beginnig of the year). She tries different models to fit the real data (all coefficients are significant)

$$\hat{Y}_t = 0.337 + 0.08 \cdot M_t \qquad R^2 = 0.89, \quad RSS = 7.0 \qquad (1)$$

$$\hat{Y}_t = -34.02 + 6.1 \cdot \log(M_t) \qquad R^2 = 0.85, \quad RSS = 8.43 \qquad (2)$$

$$\widehat{\log(Y_t)} = 0.77 + 0.0014 \cdot M_t \qquad R^2 = 0.83, \quad RSS = 0.36 \qquad (3)$$

(a) Give detailed interpretation to the models and their coefficients. For equation (3) explain why your interpretation is valid.

(a) Correct interpretation of (1) is: an increase of money supply by one million of dollars leads to the growth of GDP by 80 thousands of dollars. Correct interpretation of (2) is: an increase of money supply by one percent leads to the growth of GDP by 61 thousands of dollars on average according to the regression. Correct interpretation of (3) is: an increase of money supply by one million of dollars leads to the growth of GDP by 0.14 percent on average according to the regression.

Different alternative approaches are possible.
Theoretical approaches:

Using Calculus: Taking differentials of both sides of the equation $\ln Y_t = b_1 + b_2 M_t$ we get $\frac{dY_t}{Y_t} = b_2 dM_t$ or $\frac{dY_t}{Y_t} \cdot 100 \approx b_2 \cdot 100 dM_t$. Now taking $dM_t = 1$ we get $\frac{dY_t}{Y_t} \cdot 100 \approx b_2 \cdot 100$ So the 1 million increase in M_t leads to $b_2 \cdot 100\%$ increase of GDP.

Direct approach using function without logarithm: equation $\ln Y_t = b_1 + b_2 M_t$ is equivalent to $Y_t = e^{b_1 + b_2 M_t} = e^{b_1} e^{b_2 M_t} = A e^{b_2 M_t}$. (1)

Let M_t gets small increment dM_t so now it is $M_t + dM_t$. Substituting it into last equation we get $Y_t^{'} = A e^{b_2 M_t + dM_t}$ (2). Their ratio (dividing (2) by (1)) is $\frac{Y_t^{'}}{Y_t} = \frac{A e^{b_2 M_t + b_2 dM_t}}{A e^{b_2 M_t}} = e^{b_2 dM_t}$

Use representation $\frac{Y_t^{'}}{Y_t} = \frac{Y_t + dY_t}{Y_t} = 1 + \frac{dY_t}{Y_t}$ and Calculus formula $e^x \approx 1 + x$ for small x we get $1 + \frac{dY_t}{Y_t} \approx 1 + b_2 dM_t$, so $\frac{dY_t}{Y_t} \approx b_2 dM_t$ or $\frac{dY_t}{Y_t} \cdot 100 \approx b_2 \cdot 100 dM_t$. Now taking $dM_t = 1$ (million of dollars) we get interpretation: an increase of money supply by one million of dollars leads to the growth of GDP by $b_2 \cdot 100$ percent on average according to the regression.

Practical work with numbers is also allowed here:
Direct method using antilogarithms: let $M_t = 100$, then according to the model $\log(Y_t) = 0.77 + 0.0014 \cdot M_t$ $\ln(Y_t(100)) = 0.77 + 0.0014 \cdot 100 = 0.91$. For $M_t = 100$ $\ln(Y_t(101) = 0.77 + 0.0014 \cdot 101 = 0.9114$. The difference is $\ln(Y_t(101) - \ln(Y_t(100) = \ln\left(\frac{Y_t(100) + dY_t}{Y_t(100)}\right) = \ln\left(1 + \frac{dY_t}{Y_t(100)}\right) = 0.0014$. From here $\left(1 + \frac{dY_t}{Y_t(100)}\right) = e^{0.0014} \approx 1.00140098$. So $\frac{dY_t}{Y_t(100)} \approx 1.0014$ and percentage growth is $\frac{dY_t}{Y_t(100)} \cdot 100 = 0.14$.

Direct method using Calculus formulas for infinitesimals : From $\ln\left(1+\frac{dY_t}{Y_t(100)}\right)=0.0014$. We get $\ln\left(1+\frac{dY_t}{Y_t(100)}\right)\approx\frac{dY_t}{Y_t(100)}=0.0014$ So percentage growth is $\frac{dY_t}{Y_t(100)}\cdot 100=0.14$.

Taking differentials of both sides of the equation $\ln Y_t = b_1 + b_2 M_t$ we get $\frac{dY_t}{Y_t}=b_2 dM_t$ or taking into account that $dM_t = 1$ $\frac{dY_t}{Y_t}\cdot 100 = 0.0014\cdot 100\cdot 1 = 0.14$. So the 1 million increase in M_t leads to 0.14% increase of GDP.

(b) Why regression (1) and (3) cannot be compared directly? Explain how to use Box-Cox transformation to compare linear and logarithmic regressions. The researcher used Zarembka scaling of the linear regression instead. After this transformation new variable YZ_t was regressed on M_t with the following results

$$\hat{YZ}_t = 0.05 + 0.001\cdot M_t \qquad R^2 = 0.89, \quad RSS = 0.18 \qquad (4)$$

Explain what is Zarembka scaling, what is its aim. Show mathematically that this aim is achieved at least approximately by this procedure. *Hint: use simple calculus with some version of the second remarkable limit.*

(b) The dimension of the dependent variable in (1) and (2) is millions of dollars while dependent variable in (3) is measured in their logarithms, so their *RSS*'s also have different dimension so they are not comparable directly.

The idea of Box-Cox transformation is to represent dependent variable both in linear and logarithmic relationships in unified form $\frac{Y^\lambda - 1}{\lambda}$ with parameter λ. Changing parameter from 1 (linear relationship $\frac{Y-1}{1}=Y-1=\beta_1+\beta_2 X+u$) to the values close to zero (it is known from Calculus that $\lim_{\lambda\to 0}\frac{Y^\lambda-1}{\lambda}=\ln Y$ so if λ is small the relationship becomes approximately logarithmic $\ln Y=\beta_1+\beta_2 X$). Changing λ we can choose the relationship with smallest RSS, if it is with $\lambda\approx 1$ the linear relationship is preferred, if $\lambda\approx 0$ the model is rather logarithmic. In intermediate cases $\lambda\approx 0.5$ there is no preference and the choice can be based on some extraneous factors (interpretation, for example). The weakness of this approach is that there is no test to judge whether the difference between two functions are significant.

Alternative approach is Zarembka transformation, namely dependent variable Y is replaced by transformed one $Y^* = \frac{Y}{geometric\ mean}=\frac{Y}{(Y_1\cdot Y_2\cdot\ldots\cdot Y_n)^{\frac{1}{n}}}$. The values of Y^* in regression $Y^*=\beta_1^{(1)}+\beta_2^{(1)}X$ are directly comparable with $\ln Y=\beta_1^{(2)}+\beta_2^{(2)}X$, and the discrepancy can be estimated for significance (see Box-Cox in **(c)** below).

They are in fact comparable as $\frac{Y}{geometric\ mean}=\frac{Y}{\overline{Y}_{geom}}$ under assumption that deviations ΔY of Y from its geometric mean $\overline{Y}_{geom}$ are not 'small enough' can be represented as $\frac{Y}{\overline{Y}_{geom}}=\frac{\overline{Y}_{geom}+\Delta Y}{\overline{Y}_{geom}}=1+\alpha$, so it is only by constant equal 1 (which is absorbed by the constant term of regression) differs from α. On the other hand dependent variable $\ln\frac{Y}{\overline{Y}_{geom}}=\ln Y-\ln\overline{Y}_{geom}$ that only by constant $-\ln\overline{Y}_{geom}$ (which is also absorbed by the constant term of regression) differs from $\ln Y$. But the representation $\ln\left(\frac{Y}{\overline{Y}_{geom}}\right)=\ln(1+\alpha)\approx\alpha$ shows that $\ln Y$ and $\frac{Y}{\overline{Y}_{geom}}$ both could be represented by α and some constant, so they are directly comparable, and so residual sums of squares ***RSS*** of both regressions are comparable.

(c) How to compare regressions (1)-(3). Is it possible to say which of the models is best? Perform appropriate test(s) and explain your choice.

(c) The regressions (1) and (2) are comparable directly and judging by ***RSS*** regression (1) is better. On the other hand regressions (1) and (4) are comparable, and (4) is better. Evaluate significance of this difference. According to Box-Cox test we have to evaluate χ-square statistic and compare it with critical value of χ-square distribution with 1 degree of freedom. . $\chi^2 = (25/2)|\ln(0.36/0.18)| = 8.66 > 6.63 = \chi^2_{crit}(1\%, df = 1)$. Thus linear specification with lower RSS (after Zarembka transformation) is significantly better. As it differs only by constant term from equation (1) we choose regression (1) as best specification,

Comment: This solution requires scientific calculator to evaluate $\ln(0.36/0.18) = \ln 2$. If you don't have one it is still possible to evaluate χ-square statistic if your calculator allows to evaluate square roots. It is known from Calculus that if x is sufficiently small then $\ln(1+x) \approx x$. We have $\ln(0.36/0.18) = \ln 2 = \ln(1+1.0)$. But the number $x = 1.0$ is not small enough to use Calculus formula. So make some additional transformation

$$\ln(0.36/0.18) = \ln((\sqrt{\sqrt{\sqrt{2}}})^8) \approx 8 \cdot \ln(1.0905) = 8 \cdot \ln(1+0.0905) \approx 8 \cdot 0.0905$$

Now $(25/2) \cdot \ln(0.36/0.18) \approx 12.5 \cdot 8 \cdot 0.0905 = 9.05$ what is very close to exact value 8.66. The remaining is the same.

3. A researcher has data on *HEAL*, aggregate expenditure on health, *GNP*, aggregate gross national product, and *POP*, total population, for a sample of 70 countries in 2009. *HEAL* and *GNP* are both measured in US$ billion. *POP* is measured in million. Hypothesizing that expenditure on health per capita depends on *GNP* per capita, he fits the regression (standard errors in parentheses; *RSS* = residual sum of squares):

$$\log\frac{\widehat{HEAL}}{POP} = -3.74 + 1.29\log\frac{GNP}{POP} \qquad R^2 = 0.91 \; RSS = 14.26 \qquad (1)$$

He also runs the following regressions:

$$\log \widehat{HEAL} = -3.60 + 1.27\log GNP - 0.33\log POP \qquad R^2 = 0.95 \quad RSS = 13.90 \qquad (2)$$

$$\widehat{\log\frac{HEAL}{POP}} = -3.60 + 1.27\log\frac{GNP}{POP} - \underset{(0.04)}{0.06}\log POP \qquad R^2 = 0.91 \quad RSS = 13.90 \qquad (3)$$

(a) Give an interpretation of the slope coefficient in equation (1). Compare it with the interpretation of the slope coefficient of variable $\log GNP$ in equation (2). Give an interpretation of the coefficient of $\log POP$ in equation (2). Comment on the plausibility of coefficients paying special attention to the sign and the value of the coefficients.

Equation (1): the elasticity of expenditure on health per capita with respect to GNP per capita is 1.27. A one percent increase in GNP per capita leads to a 1.27 percent increase in health expenditure per capita. The elasticity of expenditure on health with respect to GNP, controlling for population, is 1.27. This seems plausible since health expenditure should at least keep pace with GNP. A slight excess over the unit is quite natural, since in the calculation of the partial effect of GNP growth the simultaneous growth of the population is not taken into account.
Equation (2): An increase in population, holding GNP constant, has two effects: a direct effect, which may be expected to be positive and probably with an elasticity about 1, and an indirect income effect attributable to the fact that an increase in population, holding GNP constant, means a reduction in GNP per capita. Since the income elasticity is greater than 1, the total effect should be negative, and accordingly the elasticity, -0.33, seems plausible.

(b) Demonstrate that equation (1) is a restricted version of equation (2), stating the restriction. Test the restriction, using an *F* test.
Write the first specification

$$\log\frac{HEAL}{POP} = \beta_1 + \beta_2\log\frac{GNP}{POP} + u.$$

It may be rewritten as

$$\log HEAL - \log POP = \beta_1 + \beta_2\log GNP - \beta_2\log POP + u$$

and this in turn may be rewritten

$$\log HEAL = \beta_1 + \beta_2\log GNP + (1-\beta_2)\log POP + u$$

This is a restricted version of the more general specification

$$\log HEAL = \beta_1 + \beta_2\log GNP + \beta_3\log POP + v$$

with the restriction $\beta_3 = 1 - \beta_2$.

$F(1,67) = \dfrac{(14.26-13.90)/1}{13.90/67} = 1.74$. The null hypothesis is $H_0: \beta_3 = 1-\beta_2$. The critical value of $F(1,67)$ at the 5 percent significance level is about 3.99. Hence we do not reject the restriction.

(c) How this restriction can be tested using a *t* test based on specification of the equation (3)? Perform the test using information from equation (3). Why first slope coefficient (equal to 1.27), as well as *RSS* are the same in models (2)-(3) while their R^2 coefficients are different?

Subtracting the restricted version

$$\log HEAL = \beta_1 + \beta_2 \log GNP + (1-\beta_2)\log POP + u$$

from the unrestricted version

$$\log HEAL = \beta_1 + \beta_2 \log GNP + \beta_3 \log POP + v,$$

we have

$$0 = (\beta_3 - 1 + \beta_2)\log POP$$

This means that we can rewrite the unrestricted model

$$\log\frac{HEAL}{POP} = \beta_1 + \beta_2 \log\frac{GNP}{POP} + (\beta_3 - 1 + \beta_2)\log POP + u.$$

A *t* test on the coefficient of $\log POP$ in specification (3) is thus a test of the restriction. If the coefficient is not significantly different from 0, we could drop the term and simplify the model to the restricted version. Equation (3) fits this specification. The *t* statistic is, subject to rounding error, -1.5, and hence one does not reject the restriction.

Both equations

$$\log HEAL = \beta_1 + \beta_2 \log GNP + \beta_3 \log POP + v \qquad (2)$$

and

$$\log\frac{HEAL}{POP} = \gamma_1 + \gamma_2 \log\frac{GNP}{POP} + \gamma_3 \log POP + w \qquad (3)$$

or in transformed form

$$\log HEAL = \gamma_1 + \gamma_2 \log GNP + (\gamma_3 - \gamma_2 + 1)\log POP + w$$

have the same set of dependent and independent variables with identical coefficient γ_2 of $\log GNP$ so the estimates of γ_2 should be the same (they are in fact the same equal to 1.27), also their RSS's should be the same.

On the other hand their determination coefficients $R^2 = 1 - \frac{RSS}{TSS}$ should be different as their TSS's relate to different dependent variables $\log HEAL$ and $\log\frac{HEAL}{POP}$.

SECTION B

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

4. Suppose the model $Y_i = \beta_1 + \beta_2 X_i + u_i,\ \ i = 1, \dots, n$ satisfies all assumptions of the model A. Let the values of X_i are supposed to be non-stochastic. The researcher wrongly believes that $\beta_1 = 0$ and so uses OLS estimator $\tilde{\beta}_2$ for the model $Y_i = \beta_2 X_i + u_i,\ \ i = 1, \dots, n$ instead of OLS estimator $\hat{\beta}_2 = \dfrac{\sum x_i y_i}{\sum x_i^2}$ for the model $Y_i = \beta_1 + \beta_2 X_i + u_i,\ \ i = 1, \dots, n$ where $x_i = X_i - \overline{X}$, $y_i = Y_i - \overline{Y}$.

(a) Show that OLS estimator $\tilde{\beta}_2$ for the model $Y_i = \beta_2 X_i + u_i,\ \ i = 1, \dots, n$ is $\tilde{\beta}_2 = \dfrac{\sum X_i Y_i}{\sum X_i^2}$.

a) Under assumption $\beta_1 = 0$ OLS is minimization of a function $F(\beta_2) = \sum (Y_i - \beta_2 X_i)^2$. First order condition for this is $\dfrac{d}{d\beta_2} \sum (Y_i - \beta_2 X_i)^2 = 0$ or $\sum (Y_i - \beta_2 X_i) X_i = 0$ that gives $\tilde{\beta}_2 = \dfrac{\sum X_i Y_i}{\sum X_i^2}$.

(b) Show that variance for $\tilde{\beta}_2$ is given by expression $\operatorname{var}(\tilde{\beta}_2) = \dfrac{\sigma^2}{\sum X_i^2}$.

b) $\operatorname{var}(\tilde{\beta}_2) = \operatorname{var}\left(\dfrac{\sum X_i Y_i}{\sum X_i^2} \right) = \dfrac{\operatorname{var}\left(\sum X_i Y_i\right)}{\left(\sum X_i^2\right)^2} = \dfrac{\sum \operatorname{var} X_i Y}{\left(\sum X_i^2\right)^2} = \dfrac{\sum X_i^2 (\operatorname{var} Y_i)}{\left(\sum X_i^2\right)^2} = \dfrac{\sum X_i^2 \sigma^2}{\left(\sum X_i^2\right)^2} = \dfrac{\sigma^2}{\sum X_i^2}$.

(c) If the assumption $\beta_1 = 0$ is not correct show that estimator $\tilde{\beta}_2 = \dfrac{\sum X_i Y_i}{\sum X_i^2}$ is biased and find the bias. Indicate conditions when $\tilde{\beta}_2$ is nevertheless unbiased.

c) $E\tilde{\beta}_2 = E \dfrac{\sum X_i Y_i}{\sum X_i^2} = \dfrac{\sum X_i E Y_i}{\sum X_i^2} = \dfrac{\sum X_i E(\beta_1 + \beta_2 X_i + u_i)}{\sum X_i^2} = \dfrac{\sum X_i (\beta_1 + \beta_2 X_i + E u_i)}{\sum X_i^2}$. According to G.M.C. $E u_i = 0$ so $E\tilde{\beta}_2 = \dfrac{\sum X_i (\beta_1 + \beta_2 X_i)}{\sum X_i^2} = \dfrac{\sum \beta_1 X_i + \sum \beta_2 X_i^2}{\sum X_i^2} = \beta_2 + \beta_1 \dfrac{\sum X_i}{\sum X_i^2}$. The bias is $\beta_1 \dfrac{\sum X_i}{\sum X_i^2}$. The bias could be zero if the sample is balanced so that $\sum X_i = 0$.

(d) Compare the variances of two estimators $\tilde{\beta}_2 = \dfrac{\sum X_i Y_i}{\sum X_i^2}$ and $\hat{\beta}_2 = \dfrac{\sum x_i y_i}{\sum x_i^2}$. Comment the results taking into account that assumption $\beta_1 = 0$ is not correct.

It is known that $\operatorname{var}(\hat{\beta}_2) = \dfrac{\sigma^2}{\sum x_i^2}$ and as it is known from (b) $\operatorname{var}(\tilde{\beta}_2) = \dfrac{\sigma^2}{\sum X_i^2}$ ..So we need to compare both variances. Note that $\sum x_i^2 = \sum (X_i - \overline{X})^2 = \sum X_i^2 - 2\overline{X} \sum X_i + \sum (\overline{X})^2 = \sum X_i^2 - n(\overline{X})^2$ and obviously $\sum X_i^2 - n \sum (\overline{X})^2 \le \sum X_i^2$ so $\operatorname{var}(\hat{\beta}_2) = \dfrac{\sigma^2}{\sum x_i^2} = \dfrac{\sigma^2}{\sum X_i^2 - n(\overline{X})^2} \ge \dfrac{\sigma^2}{\sum X_i^2} = \operatorname{var}(\tilde{\beta}_2)$.

That means that if assumption $\beta_1 = 0$ is not correct $\tilde{\beta}_2$ is a superefficient estimator of a slope, but at the same time it is biased.

(e) Suppose again that the assumption $\beta_1 = 0$ is wrong. Taking into account all results obtained in (a)-(c) show that under certain condition estimator $\tilde{\beta}_2$ outperforms OLS estimator $\hat{\beta}_2$ using MSE criteria. What is this condition? Pay special attention to the case of balanced sample when $\sum X_i = 0$.

(*Hint*: $MSE(b) = \mathrm{Var}(b) + [bias(b)]^2$).

e) Using MSE criteria for the comparison of the estimators we get that $\tilde{\beta}_2$ is preferable to OLS estimator $\hat{\beta}_2$ $MSE(\tilde{\beta}_2) \le MSE(\hat{\beta}_2)$. This is true if $\frac{\sigma^2}{\sum X_i^2} + \left(\beta_1 \frac{\sum X_i}{\sum X_i^2}\right)^2 \le \frac{\sigma^2}{\sum X_i^2 - n(\overline{X})^2}$.

As $\sum X_i = n\overline{X}$ it could be found from here that it is equivalent to the condition

$$\frac{\sigma^2}{\sum X_i^2} + \left(\beta_1 \frac{n\overline{X}}{\sum X_i^2}\right)^2 \le \frac{\sigma^2}{\sum X_i^2 - n(\overline{X})^2} \quad (*)$$

so $\beta_1^2 \frac{n^2(\overline{X})^2}{(\sum X_i^2)^2} \le \frac{\sigma^2}{\sum X_i^2 - n(\overline{X})^2} - \frac{\sigma^2}{\sum X_i^2} = \sigma^2 \frac{n(\overline{X})^2}{\sum X_i^2 \sum X_i^2 - n(\overline{X})^2}$

and finally $n(\beta_1)^2 \le \sigma^2 \frac{\sum X_i^2}{\sum X_i^2 - n(\overline{X})^2}$.

Note that in case $\sum X_i = n\overline{X} = 0$ the bias of $\tilde{\beta}_2$ which is $\beta_1 \frac{\sum X_i}{\sum X_i^2}$ vanishes, condition (*) collapses to equality $\frac{\sigma^2}{\sum X_i^2} = \frac{\sigma^2}{\sum X_i^2}$ so both estimators are equally efficient with population variance $\frac{\sigma^2}{\sum X_i^2}$.

5. The student runs two production function models for the same data for some developing country: $t = 1,2,...30$, where y_t is income per capita, x_t is a capital, and z_t is labor (all variables are index numbers)

$$\ln y_t = \alpha + \beta \ln x_t + 0.5 \ln z_t + v_{1t} \qquad (1)$$

$$\ln y_t = \alpha + \beta \ln x_t + \beta \ln z_t + v_{2t} \qquad (2)$$

given that x_t and z_t are deterministic sequences and $v_{1t} \sim iid(0, \sigma^2)$, $v_{2t} \sim iid(0, \sigma^2)$.

$$\ln y_t = \alpha + \beta \ln x_t + 0.5 \ln z_t + v_{1t} \qquad (1)$$

$$\ln y_t = \alpha + \beta \ln x_t + \beta \ln z_t + v_{2t} \qquad (2)$$

given that x_t and z_t are deterministic sequences and $v_{1t} \sim iid(0, \sigma^2)$, $v_{2t} \sim iid(0, \sigma^2)$.

(a) Explain how to find the least squares estimates of β. What are estimators of β for both equations (write out the explicit formulas)? What are properties of these estimators assuming equations (1) and (2) to be in turn valid models?

a) Both regressions are in fact simple linear regression models, so in both cases one should use conventional estimator of the type $\hat{\beta} = \dfrac{\mathrm{Cov}(Y, X)}{\mathrm{Var}(X)}$. But the transformations of the data needed to reduce equation to the simple regression model, are different, hence estimators will be different.

For (1) one can rewrite the model in the form $\ln y_t - 0.5 \ln z_t = \alpha + \beta \ln x_t + u_t$ and then estimate the model $y_t^{\cdot} = \alpha + \beta \ln x_t + u_t$, where $y_t^{\cdot} = \ln y_t - 0.5 \ln z_t$.

The second equation could be estimated after simple transformation $\ln y_t = \alpha + \beta(\ln x_t + \ln z_t) + v_t$, so the model is $\ln y_t = \alpha + \beta x_t^{\cdot} + u_t$ where $x_t^{\cdot} = \ln x_t + \ln z_t$.

It means that OLS gives best estimators:

for (1) we get $\hat{\beta} = \dfrac{\mathrm{Cov}(\ln y_t - 0.5 \ln z_t, \ln x_t)}{\mathrm{Var}(\ln x_t)} = \dfrac{\mathrm{Cov}(\ln y_t, \ln x_t) - 0.5\,\mathrm{Cov}(\ln z_t, \ln x_t)}{\mathrm{Var}(\ln x_t)}$

for (2) we get $\hat{\beta} = \dfrac{\mathrm{Cov}(\ln y_t, \ln x_t + \ln z_t)}{\mathrm{Var}(\ln x_t + \ln z_t)} = \dfrac{\mathrm{Cov}(\ln y_t, \ln x_t) + \mathrm{Cov}(\ln y_t, \ln z_t)}{\mathrm{Var}(\ln x_t) + \mathrm{Var}(\ln z_t) + 2\,\mathrm{Cov}(\ln x_t, \ln z_t)}$.

Both estimators of β are BLUE under assumption that models (1) and (2) are correspondingly true models as all Gauss-Markov conditions are satisfied so Gauss-Markov theorem holds.

(b) Both regressions are the restricted versions of the general model

$$\ln y_t = \alpha + \beta \ln x_t + \gamma \ln z_t + u_t \qquad (3).$$

What are the restrictions? How these restrictions could be tested **using an *F* test**

b) As it is said in the text both models are the restricted versions of the general model

$$\ln y_t = \alpha + \beta \ln x_t + \gamma \ln z_t + u_t$$

The restrictions for the models (1) and (2) are correspondingly $\gamma = 0.5$ and $\gamma = \beta$.

Any of them could be tested using *F*-test on the base of RSS or R^2 of each equation. For example, to test the restriction $\gamma = \beta$ we use F-statistics $F = \dfrac{(RSS_2 - RSS_3)/1}{RSS_3/(30-3)}$ where RSS_2 and RSS_3 are the sums of squared residuals for the models (2) and (3) (or rather $F = \dfrac{(R_3^2 - R_2^2)/1}{(1-R_3^2)/(30-3)}$). The same for the restriction $\gamma = 0.5$ (use RSS_1 and RSS_3)

(c) How restrictions in (b) could be tested using a ***t* test**?

To test the restriction $\gamma = 0.5$ for model (1) using t-test it is sufficient to evaluate t-statistic $t = \dfrac{g - 0.5}{s.e.(g)}$ where g is OLS estimator of γ in the equation (3) (test uses $df = 30 - 3$).

To test the restriction $\gamma = \beta$ for model (2) the model (3) should be transformed in the following way

$$\ln y_t = \alpha + \beta \ln x_t + \gamma \ln z_t + u_t$$

$$\ln y_t = \alpha + \beta \ln x_t - \gamma \ln x_t + \gamma \ln x_t + \gamma \ln z_t + u_t$$

$$\ln y_t = \alpha + (\beta - \gamma) \ln x_t + \gamma (\ln x_t + \ln z_t) + u_t$$

Now we use t-test to test coefficient of the variable $\ln x_t$ for significance.

(d) What is variance of each of estimators in (a)? Suppose that both restrictions in (b) are not rejected. Under what conditions the estimator of β from equation (1) is superior to that from equation (2) (and vice versa) using the criteria of minimum MSE of the estimator? Explore different cases:

- **i)** $\ln x_t$ and $\ln z_t$ are not correlated (or nearly not correlated)
- **ii)** $\ln x_t$ and $\ln z_t$ are positively correlated (perfectly or not)
- **iii)** $\ln x_t$ and $\ln z_t$ are negatively correlated (perfectly or not)

d) As both restrictions in (b) are not rejected it could be supposed that for the disturbance terms of all three models $\sigma_{v_1}^2 = \sigma_{v_2}^2 = \sigma_u^2 = \sigma^2$. Both estimators are unbiased if both restrictions in (b) are not rejected so their MSE are equal to the corresponding variances of the estimators.

The variance for the estimator of regression coefficient in the simple linear model $y_t = \beta_1 + \beta_2 x_t + u_t$ is given by expression $\sigma_{\hat{\beta}_2}^2 = \dfrac{\sigma_u^2}{n\,\mathrm{Var}(X)} = \dfrac{\sigma^2}{n\,\mathrm{Var}(X)}$.

For the model (1) (if it is valid) $\sigma_{\hat{\beta}_2}^2 = \dfrac{\sigma_{v_1}^2}{n\,\mathrm{Var}(\ln x_t)}$;

for the model (2) $\sigma_{\hat{\beta}_2}^2 = \dfrac{\sigma_{v_2}^2}{n\,\mathrm{Var}(\ln x_t + \ln z_t)} = \dfrac{\sigma^2}{n\,\mathrm{Var}(\ln x_t + \ln z_t)}$ (as the variances of the error terms are supposed to be equal).

If both models are valid, the choice of the model would depend on the comparison of two expressions for the variances $\mathrm{Var}(\ln x_t)$ and $\mathrm{Var}(\ln x_t + \ln z_t)$.

For example if $\ln x_t$ and $\ln z_t$ are not correlated (or nearly not correlated) $\mathrm{Var}(\ln x_t + \ln z_t) = \mathrm{Var}(\ln x_t) + \mathrm{Var}(\ln z_t) > \mathrm{Var}(\ln x_t)$ so the regression (2) is superior.

The same is true for the case of positive correlation $\mathrm{Var}(\ln x_t + \ln z_t) = \mathrm{Var}(\ln x_t) + \mathrm{Var}(\ln z_t) + \mathrm{Cov}(\ln x_t, \ln z_t) > \mathrm{Var}(\ln x_t)$.

For the case of negative correlation the result is uncertain as $\mathrm{Var}(\ln x_t + \ln z_t) = \mathrm{Var}(\ln x_t) + \mathrm{Var}(\ln z_t) - \mathrm{Cov}(\ln x_t, \ln z_t)$, but one could be supposed that in some cases $\mathrm{Var}(\ln x_t + \ln z_t) < \mathrm{Var}(\ln x_t)$ so the estimator based on the regression (1) can be superior.

(e) Suppose now that unrestricted form of production function $\ln y_t = \alpha + \beta \ln x_t + \gamma \ln z_t + u_t$ is valid, but the researcher realised that structural change in economy under consideration is suspected, so economic conditions in years $t = 1, 2, ..., 15$ are different from years $t = 16, 17, ..., 30$. How the presence of structural change can be tested using:

- **i)** dummy variables;
- **ii)** Chow test;

Under what condition two approaches to detecting structural change are equivalent? What approach is more informative – Chow test or dummy variables?

(e) To use dummy variables approach one should first introduce intercept dummy $D = 0$ for the period $t = 1, 2, ..., 15$ and $D = 1$ for $t = 16, 17, ..., 30$. Two different model can be evaluated

$$\ln y_t = \alpha + \delta D_t + \beta \ln x_t + \gamma \ln z_t + u_t \qquad \text{(D1)}$$

(under assumption that elasticities of income with respect to capital and labor are the same in both subperiods)
To detect structural change using (D1) one should use *t*-test for signigicance of coefficient of the variable D_t.
More advanced approach is to develop regression equation it the form

$$\ln y_t = \alpha + \delta_1 D_t + \beta \ln x_t + \delta_2 D_t \cdot \ln x_t + \gamma \ln z_t + \delta_3 D_t \cdot \ln z_t + u_t \qquad \text{(D2)}$$

(that allows elasticities of income with respect to capital and labor be different in two periods).
To detect structural change using (D2) one should use *F*-test for joint signigicance of the group of variables D_t, $D_t \cdot \ln x_t$ and $D_t \cdot \ln z_t$.

This *F*-test is equivalent to Chow test, which consists of the following steps:
1) Run equation $\ln y_t = \alpha + \beta \ln x_t + \gamma \ln z_t + u_t$ for the whole sample and get RSS_{Total}
2) Run equation $\ln y_t = \alpha + \beta \ln x_t + \gamma \ln z_t + u_t$ for subsamples $t = 1, 2, ..., 15$ and $t = 16, 17, ..., 30$ and get RSS_1 and RSS_2 correspondingly.
3) Evaluate $F = \dfrac{(RSS_{Total} - RSS_1 - RSS_2)/3}{(RSS_1 + RSS_2)/(30 - 2 \cdot 3)}$ and use critical values $F^{\alpha\%}(3, 24)$ to test whether difference between production functions for two periods is signmificant at $\alpha\%$.
F-test for the contribution of all dummies in (D2) (including all slope) is equivalent to Chow test. Dummy approach is more informative as it allows to test the significance of any group of dummies, and also allows to use one sided ***t*** tests.

The International College of Economics and Finance
Econometrics. Mid-year exam. 2016 October 27.
Suggested Solutions

IMPORTANT: Start answering each question from the new page (ask for extra paper if necessary). Structure your answers in accordance with the structure of the questions. Testing hypotheses always state clearly null and alternative hypotheses, provide critical value used for test, mentioning degrees of freedom and the significance level chosen for the test.

SECTION A

Answer **ALL** questions 1-3 from this section. (All questions from this section bear equal marks)

1. [15 marks] The researcher regresses the natural log of expenditure on videogames at 2012 prices (G_t) on the natural log of total household expenditure at 2012 prices (E_t), the natural log of the price of videogames relative to all consumer prices (PG_t), the natural log of the price of all amusements excluding videogames relative to all consumer prices (PA_t) and time t in years from $t=1$ at 2001, $t=2$ at 2002,…, till the last year in the sample 2015, with the following results

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

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) Give interpretation to the coefficients of the model. Explain mathematically that your interpretation is correct. Use confidence interval method to test whether coefficients of PG_t and PA_t are significant. Explain the meaning of confidence interval and the logic of your decision. Do your conclusions changes if it can be assumed that coefficients of PG_t and PA_t cannot be positive?

Solution: Coefficient -0.989 is the price PG elasticity of videogames expenditure G under assumption that all other variables included into equation E_t, PA_t, t are fixed (note that the sign and the value both are reasonable). Similarly, for $+1.266$ and -0.412 (in fact PG is the index of relative price of videogames, but it is not expected to be mentioned here). The coefficient $+0.004$ of time variable t shows that the growth rate of expenditure on videogames is 4% per year under assumption that all other explanatory variables E_t, PG, PA_t do not change.

The interpretation of coefficients as elasticities follows from the following: if $\ln G = \beta_1 + \beta_2 \ln E + ...$ then $\frac{\partial G}{G} = \beta_2 \frac{\partial E}{E}$ keeping other variables constant. So $\beta_2 = \frac{\partial G}{G} \cdot 100\% : \frac{\partial E}{E} \cdot 100\%$.

If $\ln G = \beta_1 + \beta_5 t + ...$ then $\frac{\partial G}{G} = \beta_5$ keeping other variables constant. So $\frac{\Delta G}{G} \cdot 100\% = \beta_5 \cdot 100\%$.

Alternatively If $\ln G(t) = \beta_1 + \beta_5 t + ...$ and $\ln G(t+1) = \beta_1 + \beta_5 (t+1) + ...$ then keeping other variables constant. $\ln G(t+1) - \ln G(t) = \beta_5$ and so $\ln \frac{G(t+1)}{G(t)} = \beta_5$ or $\frac{G(t+1)}{G(t)} = e^{\beta_5}$ and $\frac{G(t+1)}{G(t)} - 1 = \frac{G(t+1) - G(t)}{G(t)} = e^{\beta_5} - 1 \approx \beta_5$. From here $\frac{G(t+1) - G(t)}{G(t)} \cdot 100\% \approx \beta_5 \cdot 100\%$.

The confidence interval gives an interval in which we are 95% sure of containing the true value of the parameter. It constitutes of all null hypotheses compatible with the data. As it includes zero hypothesis $H_0 : \beta_{PG_t} = 0$ is not rejected.

The 95% confidence interval for β_{PG_t} under (15 – 5 = 10 degrees of freedom) is $-0.989 \pm 2.228 \times 0.496$ - $0 \in [-1.982688; 0.004688]$ ⇔ insignificant.
For β_{PA_t} we get $-0.412 \pm 2.228 \times 0.144$ - $0 \notin [-0.821952; -0.091168]$ (but $-0.412 + 3.169 \times 0.144 = 0.044$ >0) - significant only at 5%.

If coefficients of pb_t and pa_t cannot be positive we can use one-sided values 1.812 instead of 2.228 (5%) and 2.764 instead of 3.169. So for $H_0 : \beta_{PG_t} = 0$, $H_a : \beta_{PG_t} < 0$ we get $t = \frac{-0.989}{0.446} = 2.217 > 1.812$ (significant), and for $H_0 : \beta_{PA_t} = 0$, $H_a : \beta_{PA_t} < 0$ we get $t = \frac{-0.412}{0.144} = 2.8611 > 2.764$ - now significant at 1%.

(b) The researcher suspects that the theoretical coefficients of PG_t and PA_t are equal in value. How to test this hypothesis against alternative hypothesis that they are not equal? What additional information is needed for this? Is it possible to test this hypothesis against the alternative that the coefficient of PG_t is smaller than the coefficient of PA_t? What is economic meaning of these hypotheses?

Solution. First let us wright theoretical equation, corresponding estimated equation above

$$G_t = \beta_1 + \beta_2 E_t + \beta_3 PG_t + \beta_4 PA_t + \beta_5 t + u_t$$

The restriction is $\beta_3 = \beta_4$ with the meaning that price elasticities of videogames consumption and consumption of other amusements are equal.
To test this restriction the researcher needs to run equation

$$G_t = \beta_1 + \beta_2 E_t + \beta_3 (PG_t + PA_t) + \beta_5 t + u_t$$

memorize the R_R^2 of restricted equation and compare it with $R_U^2 = 0.906$ of initial unrestricted equation using F-test with test statistic $F = \frac{R_U^2 - R_R^2}{(1 - R_U^2)/(15-5)}$. If $F > F_{crit}$ the null hypothesis $H_0 : \beta_3 = \beta_4$ is rejected in favor of alternative $H_a : \beta_3 \neq \beta_4$.
Under one sided alternative $H_a : \beta_3 < \beta_4$ the F-test is useless. Here we should use reparametrization of the original equation

$$G_t = \beta_1 + \beta_2 E_t + \beta_3 PG_t + \beta_4 PA_t + \beta_5 t + u_t$$

as

$$G_t = \beta_1 + \beta_2 E_t + (\beta_3 - \beta_4) PG_t + \beta_4 (PG_t + PA_t) + \beta_5 t + u_t$$

or

$$G_t = \beta_1 + \beta_2 E_t + \beta \cdot PG_t + \beta_4 (PG_t + PA_t) + \beta_5 t + u_t \quad \text{where } \beta = \beta_3 - \beta_4$$

After running last equation we use une sided t-test for coefficient β: $H_0 : \beta = 0$ versus $H_0 : \beta < 0$.
If PG and PA elasticities are equal it means that the sensitivity of videogames consumption to the changes in relative prices is the same as the sensitivity to the changes in relative prices on other amusements. One-sided alternative hypothesis means that sensitivity to the changes in relative prices on videogames is greater in absolute value (both elasticities being negative).

2. [15 marks] Consider a simple linear regression model:

$$Y_i = \beta_1 + \beta_2 X_i + u_i\,,\ i = 1, 2, ..., n \qquad (1)$$

where $\mathrm{E}(u_i) = 0$; $\mathrm{E}(u_i^2) = \sigma^2$ and $\mathrm{E}(u_i u_j) = 0$ if $i \neq j$, X_i assumed to be non-stochastic.

(a) Prove that the OLS estimator of β_2 is unbiased. Explain how the assumptions of the model relate to the conditions $\mathrm{var}(u_i) = \sigma^2$ and $\mathrm{cov}(u_i, u_j) = 0, i \neq j$. Discuss efficiency of the OLS estimator of β_2 (no proof is expected). Assume now that the volume of the sampe n remaining finite, tends to infinity, so we can now discuss consistency of the estimators. Is the OLS estimator of β_2 consistent? Explain in detail. What difference would it make if it could be assumed that $\beta_1 = 0$?

Solution. OLS estimator of OLS is $\hat{\beta}_2^{OLS} = \dfrac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})_i^2} = \beta_2 + \dfrac{\sum (X_i - \overline{X})u_i}{\sum (X_i - \overline{X})_i^2}$, so

$E(\hat{\beta}_2^{OLS}) = \beta_2 + \dfrac{\sum (X_i - \overline{X})Eu_i}{\sum (X_i - \overline{X})_i^2} = \beta_2$ (this proves that estimator $\hat{\beta}_2^{OLS} = \dfrac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})_i^2}$ is unbiased).

$\mathrm{var}(u_i) = E(u_i)^2 - (Eu_i)^2 = E(u_i)^2 - 0 = \sigma^2$

$i \neq j, \mathrm{cov}(u_i, u_j) = E(u_i u_j) - E(u_i)E(u_j) = 0 - 0 = 0$

Let $a_i = \dfrac{X_i - \overline{X}}{\sum \left(X_i - \overline{X}\right)^2}$. As $\hat{\beta}_2 = \beta_2 + \sum_{i=1}^{n} a_i u_i$ then

Direct approach

$$\mathrm{var}(\hat{\beta}_2) = \mathrm{var}(\sum_{i=1}^{n} a_i u_i) = \sum_{i=1}^{n} a_i^2 \,\mathrm{var}\, u_i + \sum_{i=1}^{n} \sum_{j \neq i} a_i a_j \,\mathrm{cov}(u_i; u_j) = \sum_{i=1}^{n} a_i^2 \,\mathrm{var}\, u_i + 0 = \sigma_u \sum_{i=1}^{n} a_i^2 = \frac{\sigma_u^2}{\sum_{j=1}^{n} (X_j - \overline{X})^2}$$

as $\sum_{i=1}^{n} a_i^2 = \dfrac{1}{\sum_{j=1}^{n} (X_j - \overline{X})^2}$.

Alternative approach

As $\hat{\beta}_2 = \beta_2 + \sum_{i=1}^{n} a_i u_i$ and $\hat{\beta}_2^{OLS}$ is unbiased then $\sigma_{\hat{\beta}_2}^2 = E\left\{(\hat{\beta}_2 - E(\hat{\beta}_2))^2\right\} = E(\hat{\beta}_2 - \beta_2)^2 = E\left\{\left(\sum_{i=1}^{n} a_i u_i\right)^2\right\} =$

$$= E\left\{\sum_{i=1}^{n} a_i^2 u_i^2 + \sum_{i=1}^{n} \sum_{j \neq i} a_i a_j u_i u_j\right\} = \sum_{i=1}^{n} a_i^2 E(u_i^2) + \sum_{i=1}^{n} \sum_{j \neq i} a_i a_j E(u_i u_j) = \sum_{i=1}^{n} a_i^2 \sigma_u^2 = \sigma_u^2 \sum_{i=1}^{n} a_i^2 = \frac{\sigma_u^2}{\sum_{j=1}^{n} (X_j - \overline{X})^2}$$

$\mathrm{var}(\hat{\beta}_2) = \dfrac{\sigma_u^2}{\sum_{j=1}^{n} (X_j - \overline{X})^2} = \dfrac{\sigma_u^2}{n\,\mathrm{Var}(X)}$. It follows from here that $\lim_{n \to \infty} V(\hat{\beta}_2^{OLS}) \to 0$ since as the sample size increases $\sum (X_i - \overline{X})^2 = n\,\mathrm{Var}(X)$ must increase. Hence a sufficient condition for consistency holds $\Longrightarrow$ $\hat{\beta}_2^{OLS}$ is a consistent estimator of β_2.

As assumptions of Gauss-Markov theorem hold it states that OLS estimator $\hat{\beta}_2^{OLS}$ is the efficient estimator.

The situation changes if it could be assumed that $\beta_1 = 0$: now OLS estimator is $\hat{\beta}_2^{OLS} = \frac{\sum X_i Y_i}{\sum X_{ii}^2}$ so the estimator $\frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{\sum (X_i - \overline{X})_i^2}$ becomes inefficient. But it is still unbiased and consistent.

Alternative approach to the proof of consistency of OLS estimator.

If $\hat{\theta}$, based on a sample of size T, is a consistent estimator of θ then $\Pr(|\hat{\theta} - \theta| > e) \to 0$ as $T \to \infty$ for every $e > 0$. Another way of expressing this is that $\hat{\theta}$ converges in probability to θ. In short, we can write the above statement as $\operatorname{plim} \hat{\theta} = \theta$, where plim stands for the probability limit. Hence if $\operatorname{plim} \hat{\theta} = \theta$ then $\hat{\theta}$ is a consistent estimator of θ.

The sufficient condition for consistency is $E(\hat{\theta}) = \theta$ or $\lim_{T \to \infty} E(\hat{\theta}) = \theta$ and $\lim_{T \to \infty} \operatorname{Var}(\hat{\theta}) = \theta$.

Let the model be $y_t = \beta x_t + u_t$, $t = 1, 2, \ldots, T$, $E(u_t) = 0$, $E(u_t^2) = \sigma^2$ and $E(u_s u_t) = 0$ if $s \neq t$, for all $s, t = 1, 2, \ldots, T$.

$$\hat{\beta} = \frac{\sum x_t y_t}{\sum x_t^2} = \beta + \frac{\sum x_t u_t}{\sum x_t^2}$$

$$\operatorname{plim}(\hat{\beta}) = \beta + \frac{\operatorname{plim}(\sum x_t u_t)/T}{\operatorname{plim}(\sum x_t^2)/T} = \beta + \frac{\operatorname{cov}(x, u)}{\operatorname{var}(x)} = \beta + \frac{0}{\sigma_x^2} = \beta .$$

Hence $\hat{\beta}$ is a consistent estimator of β.

Assumptions: The xs are non-stochastic, or $\operatorname{cov}(x, u) = 0$.

(b) Let $\overline{X}$ and $\overline{Y}$ are sample means of X_i and Y_i. Show that fitted line $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_i$ (where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the OLS estimators of β_1 and β_2) must pass through the point $(\overline{X}, \overline{Y})$.

Let the relationship between the variables X and Y is described by the linear model specified above $Y_i = \beta_1 + \beta_2 X_i + u_i$ ($\beta_1 \neq 0$). Given a sample of n observations, consider alternative estimator of β_2: $\tilde{\beta}_2 = \frac{\overline{Y}}{\overline{X}}$ (the average value of Y divided by the average value of X). Discuss whether this estimator is unbiased, efficient, and consistent. How these properties would change under assumption that $\beta_1 = 0$?

Solution: One of two least square equations is $\sum Y_i = \hat{\beta}_1 + \hat{\beta}_2 \sum X_i$, dividing it by n we get $\overline{Y} = \hat{\beta}_1 + \hat{\beta}_2 \overline{X}$, so the point $(\overline{Y}, \overline{X})$ lies on the fitted regression line $\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_i$.

Since

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$\overline{Y} = \beta_1 + \beta_2 \overline{X} + \overline{u}$$

and

$$\tilde{\beta}_2 = \frac{\overline{Y}}{\overline{X}} = \frac{\beta_1 + \beta_2 \overline{X} + \overline{u}}{\overline{X}} = \frac{\beta_1}{\overline{X}} + \beta_2 + \frac{\overline{u}}{\overline{X}} .$$

Hence, assuming that X is non-stochastic,

$$E(\tilde{\beta}_2) = \frac{\beta_1}{\overline{X}} + \beta_2 + \frac{1}{\overline{X}} E(\bar{u}) = \frac{\beta_1}{\overline{X}} + \beta_2$$

since $E(\bar{u}) = 0$. Thus $\tilde{\beta}_2$ is biased unless $\beta_1 = 0$, and the direction of the bias depends on the sign of both β_1 and $\overline{X}$.

Since $\underset{n\to\infty}{\text{plim}} E(\tilde{\beta}_2) \neq \beta_2$, the estimator is not consistent.

Special case: $\beta_1 = 0$. $\tilde{\beta}_2$ is now unbiased and $\text{var}(\tilde{\beta}_2) = \text{var}(\frac{\overline{Y}}{\overline{X}}) = \text{var}(\frac{\frac{1}{n}\sum_{j=1}^{n} Y_i}{\overline{X}}) = \frac{1}{n^2} \frac{n\sigma^2}{(\overline{X})^2} = \frac{\sigma^2}{n(\overline{X})^2} \to 0$ as $n \to \infty$. Hence if $\beta_0 = 0$, then $\tilde{\beta}_2$ is a consistent estimator of β_2.

In comparison to the OLS estimator $\beta_1^{\hat{OLS}} = \frac{\sum X_i Y_i}{\sum X_i^2}$, $\tilde{\beta}_2$ is inefficient. If the Gauss-Markov assumptions hold, the OLS estimator is the most efficient estimator.

3. [15 marks]. The researcher is interested in testing of the hypothesis that gender of schoolchildren influence the consumption of unhealthy food (burgers, chips, cola and so on) in the school cafeteria. She collected data on the number of cases of buying unhealthy food per month by a respondent schoolchild, variable U_i. Dummy variable B_i is equal to 1 if respondent is a boy and 0 if respondent is a girl. Dummy variable G_i is equal to 1 if respondent is a girl and 0 if respondent is a boy. Consider three models

$$U_i = \gamma_1 + \gamma_2 G_i + \varepsilon_i, \quad i = 1, \dots, n \quad (1),$$

$$U_i = \beta_1 + \beta_2 B_i + \delta_i, \quad i = 1, \dots, n \quad (2)$$

and

$$U_i = \alpha_1 B_i + \alpha_2 G_i + u_i, \quad i = 1, \dots, n \quad (3)$$

where $\mathrm{E}(u_i) = 0$; $\mathrm{E}(u_i^2) = \sigma^2$ and $\mathrm{E}(u_i u_j) = 0$ if $i \neq j$, similar conditions hold for ε_i and δ_i.

(a) Explain why the coefficients of these models should satisfy the conditions $\beta_2 = -\gamma_2$, $\beta_1 = \gamma_1 + \gamma_2$, $\gamma_1 = \beta_1 + \beta_2$, $\gamma_1 = \alpha_1$, $\gamma_2 = \alpha_2 - \alpha_1$.

Solution based on the meaning of parameters. For equation (1) reference category is 'boy', so γ_1 shows the level of consumption of unhealthy food consumed by a boy, and γ_2 represent the 'premium' (positive or negative) – additional amount of unhealty food consumed by a girl so $\gamma_1 + \gamma_2$ is total amount of unhealthy food consumed by a girl (on average). Vice versa β_1 shows the level of consumption of unhealthy food consumed by a girl (reference category for equation (2), and β_2 represent the 'premium' (positive or negative) – additional amount of unhealty food consumed by a boy so $\beta_1 + \beta_2$ is the total amount of unhealthy food consumed by a boy (on average). From here we get $\beta_2 = -\gamma_2$, $\beta_1 = \gamma_1 + \gamma_2$ and $\gamma_1 = \beta_1 + \beta_2$. The same method for other equalities.

Brief formal solution.

$$U_i = \gamma_1 + \gamma_2 G_i + \varepsilon_i = \gamma_1 + \gamma_2 (1 - B_i) + \varepsilon_i = (\gamma_1 + \gamma_2) - \gamma_2 B_i + \varepsilon_i$$

$$U_i = \gamma_1 + \gamma_2 G_i + \varepsilon_i = \gamma_1 (1 + B_i - B_i) + \gamma_2 G_i + \varepsilon_i = \gamma_1 B_i + \gamma_1 (1 - B_i) + \gamma_2 G_i + \varepsilon_i = \gamma_1 B_i + (\gamma_1 + \gamma_2) G_i + \varepsilon_i$$

Thus all three models (1), (2), and (3) are in fact identical. From here we get all equalities in a).

Alternative solution. We will show this method on the last two equalities.
Now set simultaneously $G_i = 0$ (and so $B_i = 1$) in equations (1) and (3): $U_i = \gamma_1 + 0 + \varepsilon_i$, $U_i = \alpha_1 1 + 0 + u_i$, that gives $\alpha_1 + u_i = \gamma_1 + \varepsilon_i$ and $E(\alpha_1 + u_i) = \alpha_1 + Eu_i = \alpha_1 = E(\gamma_1 + \varepsilon_i) = \gamma_1 + E\varepsilon_i = \gamma_1$. Setting instead $G_i = 1$ (and so $B_i = 0$) in equations (1) and (3) gives $U_i = \gamma_1 + \gamma_2 1 + \varepsilon_i = \alpha_1 + \gamma_2 1 + \varepsilon_i$, and $U_i = \alpha_1 0 + \alpha_2 1 + u_i$, so, using the same method, $\alpha_1 + \gamma_2 = \alpha_2$ or $\gamma_2 = \alpha_2 - \alpha_1$.

(b) Derive OLS estimators for the coefficients of these models and show that estimated coefficients satisfy the corresponding conditions $\hat{\beta}_2 = -\hat{\gamma}_2$, $\hat{\beta}_1 = \hat{\gamma}_1 + \hat{\gamma}_2$, $\hat{\gamma}_1 = \hat{\beta}_1 + \hat{\beta}_2$, $\hat{\gamma}_1 = \hat{\alpha}_1$, $\hat{\gamma}_2 = \hat{\alpha}_2 - \hat{\alpha}_1$. Which of three models (1), (2), (3) have equal R^2 after using OLS for their estimation? Show that testing hypothesis $\gamma_2 = 0$ for the model (1), is equivalent to testing hypothesis $\alpha_1 = \alpha_2$ for the model (3).

Solution. Let us check this for OLS estimators.
Estimate coefficients for model (1). Assume without any loss of generality that first b observations correspond to boys and next g observations correspond to girls $(b + g = n)$.

Thus: $\sum_{i=1}^{n} (U_i - \hat{\gamma}_1 - \hat{\gamma}_2 G_i)^2 \to \min_{\hat{\gamma}_1, \hat{\gamma}_2}$

$$\sum_{i=1}^{n}(U_i-\hat{\gamma}_1-\hat{\gamma}_2 G_i)^2=\sum_{i=1}^{b}(U_i-\hat{\gamma}_1)^2+\sum_{i=1}^{g}(U_i-\hat{\gamma}_1-\hat{\gamma}_2)^2 \text{ (for boys } G_i=0\text{)}$$

We write the first order conditions (F.O.C.)

$$\begin{cases}-2\sum_{i=1}^{b}(U_i-\hat{\gamma}_1)-2\sum_{i=1}^{g}(U_i-\hat{\gamma}_1-\hat{\gamma}_2)=0\\ -2\sum_{i=1}^{g}(U_i-\hat{\gamma}_1-\hat{\gamma}_2)=0\end{cases}$$

Substituting from the second equation into the first $\sum_{i=1}^{b}(U_i-\hat{\gamma}_1)=0$ and $\hat{\gamma}_1=\overline{U}_b$

From the second equation: $\sum_{i=1}^{g}(U_i-\hat{\gamma}_1-\hat{\gamma}_2)=0$, so $\sum_{i=1}^{g}(U_i-\overline{U}_b-\hat{\gamma}_2)=0$ and $\hat{\gamma}_2=\overline{U}_g-\overline{U}_b$

Second order conditions (S.O.C.) are trivial here.

Doing the same with the second equation (or just by analogy) we get

$\hat{\beta}_1=\overline{U}_g$ and $\hat{\beta}_2=\overline{U}_b-\overline{U}_g$

Comparing them with the previous results

$\hat{\gamma}_1=\overline{U}_b$ and $\hat{\gamma}_2=\overline{U}_g-\overline{U}_b$

we can see that $\hat{\beta}_2=-\hat{\gamma}_2$ and $\hat{\beta}_1=\hat{\gamma}_1+\hat{\gamma}_2$

For the model (3)

$$\sum_{i=1}^{n}(U_i-\hat{\alpha}_1 B_i-\hat{\alpha}_2 G_i)^2\to\min_{\hat{\alpha}_1,\hat{\alpha}_2}$$

$$\sum_{i=1}^{n}(U_i-\hat{\alpha}_1 B_i-\hat{\alpha}_2 G_i)^2=\sum_{i=1}^{b}(U_i-\hat{\alpha}_1)^2+\sum_{i=1}^{g}(U-\hat{\alpha}_2)^2$$

$$\hat{\alpha}_1=\frac{1}{b}\sum_{i=1}^{b}U_i=\overline{U}_b$$

$$\hat{\alpha}_2=\frac{1}{g}\sum_{i=1}^{g}U_i=\overline{U}_g$$

S.O.C. is trivial.

Thus: $\hat{\gamma}_1=\hat{\alpha}_1$, $\hat{\gamma}_2=\hat{\alpha}_2-\hat{\alpha}_1$.

From equalities (*) and (**) we also can conclude that all three estimated models should have the same R^2.

Now consider t-test. We have shown in b), that $\hat{\gamma}_2=\hat{\alpha}_2-\hat{\alpha}_1$

Thus $\mathrm{Var}(\hat{\gamma}_2)=\mathrm{Var}(\hat{\alpha}_2-\hat{\alpha}_1)$, because data and formulas for righthand and lefthand parts are the same. Thus *t*-statistics will be the same (equal degrees of freedom, because we estimate two parameters in both cases).

Alternatively, if we substitute $B_i=1-G_i$ into (3):

$$U_i=\alpha_1(1-G_i)+\beta_2 G_i+u_i=\alpha_1+(\alpha_2-\alpha_1)G_i+u_i$$

Thus $\gamma_1=\alpha_1$, $\gamma_2=\alpha_2-\alpha_1$. Thus null and alternative hypotheses are the same in both cases, *t*-statistics are the same. So the tests under consideration are equivalent.

SECTION B

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

4. [15 marks]. Consider the model $Y_i = \beta X_i + u_i, \quad i = 1, \ldots, n$ where $Eu_i = 0$, $Eu_i^2 = \sigma^2$, and $E(u_i u_j) = 0, i \neq j$, and X_i are assumed to be non-stochastic.

(a) Show that OLS estimator of β is $\hat{\beta} = \dfrac{\sum X_i Y_i}{\sum X_i^2}$.

Solution. OLS: $F(\beta_2) = \sum (Y_i - \beta_2 X_i)^2 \to \min$. F.O.C.: $\dfrac{d}{d\beta_2} \sum (Y_i - \beta_2 X_i)^2 = -2\sum (Y_i - \beta_2 X_i) = 0$ or $\sum (Y_i - \beta_2 X_i) X_i = 0 \Rightarrow \tilde{\beta}_2 = \dfrac{\sum X_i Y_i}{\sum X_i^2}$

SOC: $\dfrac{d^2}{d\beta_2^2} \sum (Y_i - \beta_2 X_i)^2 = \dfrac{d}{d\beta_2}(-2\sum (Y_i - \beta_2 X_i) X_i) = 2\sum X_i^2 > 0 \Leftrightarrow \min$

(b) What is variance of $\hat{\beta} = \dfrac{\sum X_i Y_i}{\sum X_i^2}$?

Solution. $\operatorname{var}(\hat{\beta}) = \operatorname{var}\left(\dfrac{\sum X_i Y_i}{\sum X_i^2}\right) = \dfrac{\operatorname{var}\left(\sum X_i Y_i\right)}{\left(\sum X_i^2\right)^2} = \dfrac{\operatorname{var}\left(\sum X_i u_i\right)}{\left(\sum X_i^2\right)^2} = \dfrac{\left(\sum X_i^2 \operatorname{var} u_i + \sum_{i \neq j} X_i X_j \operatorname{cov}(u_i; u_j)\right)}{\left(\sum X_i^2\right)^2} =$

$= \sigma^2 \dfrac{\sum X_i^2}{\left(\sum X_i^2\right)^2} = \dfrac{\sigma^2}{\sum X_i^2}$

(c) Let $c_i = \dfrac{X_i}{\sum X_i^2}$. Which of the following equalities are true and under what condition(s) each of them is true: 1) $\sum c_i = 0$; 2) $\sum c_i^2 = \dfrac{1}{\sum X_i^2}$, 3) $\sum c_i X_i = 1$; 4) $\sum c_i \hat{Y}_i = \hat{\beta}$; 5) $\sum c_i Y_i = \hat{\beta}$.

Solution.

1) $\sum c_i = \sum \dfrac{X_i}{\sum X_i^2} = \dfrac{\sum X_i}{\sum X_i^2} = 0 \Leftrightarrow \overline{X} = 0$.

2) $\sum c_i^2 = \sum \dfrac{X_i^2}{(\sum X_i^2)^2} = \dfrac{\sum X_i^2}{(\sum X_i^2)^2} = \dfrac{1}{\sum X_i^2}$

3) $\sum c_i X_i = \sum \dfrac{X_i}{\sum X_i^2} X_i = \dfrac{\sum X_i^2}{\sum X_i^2} = 1$

4) $\sum c_i \hat{Y}_i = \sum \dfrac{X_i}{\sum X_i^2} \hat{Y}_i = \dfrac{\sum X_i \hat{\beta} X_i}{\sum X_i^2} = \hat{\beta} \dfrac{\sum X_i^2}{\sum X_i^2} = \hat{\beta}$

5) $\sum c_i Y_i = \sum \dfrac{X_i}{\sum X_i^2} Y_i = \dfrac{\sum X_i Y_i}{\sum X_i^2} = \hat{\beta}$ as it is OLS estimator.

(d) Let $\tilde{\beta} = \sum g_i Y_i$ be any linear estimator of β. Under what condition this estimator is unbiased? Does $\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2}$ belong to this class of estimators? Is it unbiased?

Solution. $E\tilde{\beta}_2 = E(\sum g_i Y_i) = E(\sum g_i \beta X_i + \sum g_i u_i) = \beta \sum g_i X_i + \sum g_i E u_i = \beta \sum g_i X_i + 0 = \beta \Leftrightarrow \sum g_i X_i = 1$.

Of course $\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2} = \sum c_i Y_i$ is also linear and $\sum c_i X_i = 1$ (property 3 in **c)**), so it is unbiased.

(e) Let $h_i = g_i - c_i$ where at least one of h_i is not equal to zero, and let $\tilde{\beta} = \sum g_i Y_i$ be any linear unbiased estimator of β; $\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2}$. Show that $\operatorname{var}(\hat{\beta}) < \operatorname{var}(\tilde{\beta})$ and so the estimator $\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2}$ is BLUE (Best Linear Unbiased Estimator) (*Gauss-Markov Theorem for the regression without constant term*).

Solution. It is known from **b)** that $\operatorname{var}(\hat{\beta}) = \frac{\sigma_u^2}{\sum X_i^2}$.

From $h_i = g_i - c_i$ $g_i = c_i + h_i$. For estimator $\tilde{\beta} = \sum g_i Y_i$ $\sum g_i Y_i = \beta \sum g_i X_i + \sum g_i u_i$. If estimator $\tilde{\beta} = \sum g_i Y_i$ is unbiased then $\sum g_i X_i = 1$ (see **d)**), so $\tilde{\beta} = \sum g_i Y_i = \beta + \sum g_i u_i$. From here $\tilde{\beta} - \beta = \sum g_i u_i$. Let us evaluate the variance of estimator $\tilde{\beta}$. As it is unbiased $\operatorname{var}(\tilde{\beta}) = E(\tilde{\beta} - E(\tilde{\beta}))^2 = E(\tilde{\beta} - \beta)^2 = E(\sum g_i u_i)^2$. Now $\operatorname{var}(\tilde{\beta}) = E(\sum g_i u_i)^2 = E(\sum (c_i u_i + h_i u_i)^2 = \sum c_i^2 E u_i^2 + \sum h_i^2 E u_i^2 + 2\sum_{i \neq j} c_i h_i E(u_i u_j)$. The last sum is zero as $E(u_i u_j) = 0, i \neq j$. As $E u_i^2 = \sigma^2$ and $\sum c_i^2 = \frac{1}{\sum X_i^2}$ (property 2 from **c)**) we have $\operatorname{var}(\tilde{\beta}) = \sigma^2(\sum c_i^2 + \sum h_i^2) = \frac{\sigma^2}{\sum X_i^2} + \sigma^2 \sum h_i^2 = \operatorname{var}(\hat{\beta}) + \sigma^2 \sum h_i^2$. Thus inequality $\operatorname{var}(\hat{\beta}) < \operatorname{var}(\tilde{\beta})$ is proved.

5. A student explores predictive properties of regression models using data on natural logarithm of expenditures on housing H_t , natural logarithm of disposable personal income Y_t (both in billions of dollars) and natural logarithm of relative prices index on housing P_t in USA for 1979-2014.

(a) First she runs a simple regression model for the sample period 1979-2010

$$\begin{aligned} H_t &= 20.59 - 3.20P_t \qquad R^2 = 0.16 \\ &\quad (6.06) \quad (1.33) \qquad RSS = 3.30 \end{aligned} \qquad (1)$$

(here and further standard errors in parenthesis), and then she obtains predictions H_{T+p} for 2011-2014 ($T = 2010,\ p = 1, 2, 3, 4$) using equation (1): $\hat{H}_{T+p} = \hat{\beta}_1 + \hat{\beta}_2 P_{T+p}$. What is the differrence between forecast and prediction. Show that $E\hat{H}_{T+p} = H_{T+p}$. What is the variance of the prediction error $f_{T+p} = H_{T+p} - H_{T+p}$? Explain how different factors influence the value of the variance of the prediction error.

Solution. If you are willing to predict a particular value of H_{T+p}, without knowing the actual value of P_{T+p}, you are said to be making a forecast, if you take into account the actual values of P_{T+p} you are said to be making a prediction. Forecasts are less accurate than predictions because they are subject to an additional source of error, the error in the prediction of P_{T+p}.

Property $E\hat{H}_{T+p} = H_{T+p}$ is equivalent to $Ef_{T+p} = E(H_{T+p} - H_{T+p}) = 0$. Really

$$\begin{aligned} E(f_{T+p}) &= E(\hat{H}_{T+p}) - E(H_{T+p}) \\ &= E(\hat{\beta}_1 + \hat{\beta}_2 P_{T+p}) - E(\beta_1 + \beta_2 P_{T+p} + u_{T+p}) \\ &= E(\hat{\beta}_1) + P_{T+p} E(\hat{\beta}_2) - \beta_1 - \beta_2 P_{T+p} - E(u_{T+p}) \\ &= \beta_1 + \beta_2 P_{T+p} - \beta_1 - \beta_2 P_{T+p} = 0 \end{aligned}$$

In the simple regression case, the population variance of f_{T+p} is given by

$$\sigma^2_{f_{T+p}} = \left\{ 1 + \frac{1}{n} + \frac{(P_{T+p} - \overline{P})^2}{n\mathrm{Var}(P)} \right\} \sigma_u^2$$

where $\overline{P}$ and $\mathrm{Var}(P)$ are the sample period mean and variance of P. It represents minimum variance from all linear unbiased estimators of f_{T+p}. The expression for the variance implies that the farther is the value of P from its sample mean, the larger will be the population variance of the prediction error. It also implies, again unsurprisingly, that, the larger is the sample, the smaller will be the population variance of the prediction error, with a lower limit of σ_u^2. As the sample becomes large, *a* and *b* will tend to their true values (provided that the Gauss-Markov conditions hold), so the only source of error in the prediction will be u_{T+p}, and by definition this has population variance σ_u^2.

(b) Then she runs multiple regression for the period 1979-2010

$$\begin{aligned} H_t &= -1.96 + 1.13Y_t - 0.24P_t \qquad R^2 = 0.99 \\ &\quad (0.39)\ (0.01)\ \ (0.08) \qquad RSS_T = 0.008958 \end{aligned} \qquad (2)$$

(the same equation for the total period (years 1979-2014) gives $RSS_{T+P} = 0.009218$).

Unfortunately explicit expression for the vaiance of the prediction error needs matrix algebra, but there is a simple way to get standard errors for each observation: one can introduce four dummy variables $D1 = 1$ only for the year 2011 and zero for other years, $D2 = 1$ only for the year 2012, and so $D3$ and $D4$. Adding them into equation (2) the student obtains both the values of prediction error for each year of the prediction period and their standard errors:

$$H_t = -1.96 + 1.13Y_t - 0.24P_t + 0.016D1 + 0.002D2 + 0.003D3 + 0.003D4 \qquad R^2 = 0.99 \tag{3}$$
$$(0.39)\ (0.01)\ \ (0.08)\ \ \ (0.019)\ \ \ (0.019)\ \ \ (0.019)\ \ \ (0.019)\ \ RSS = 0.008958$$

Are coefficients of dummy variables significant? Is the group of all dummy variables $D1, D2, D3, D4$ significant? What is your conclusion on the quality of the prediction? Why all coefficients of non-dummy variables, their standard errors, the values of R^2 and RSS are identical in equations (2) and (3)?

Solution. All dummy variables $D1, D2, D3, D4$ are insignificant (for example $t(D1) = \frac{0.016}{0.019} < 1$ and so on). To test whether they are significant together we have to test $H_0: \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$ against $H_a:$ at least one of $\beta_4, \beta_5, \beta_6, \beta_7$ is not zero, using F-test for four simultaneous restrictions. $F = \frac{(RSS_R - RSS_U)/(df(u) - df(R))}{RSS_U/(df(u)} = \frac{(0.009218 - 0.008958)/4}{0.008958/(36-7)} = 0.21$ which is insignificant as $F(crit.,5\%,4,29) = 2.70$.

The coefficients of dummy variables are the values of prediction errors, so the prediction errors taken together are insignificant, so the prediction works well.

One do not even have to run the regression with the dummy variables to perform the test, because RSS^D_{T+P} is identical to RSS_T, the sum of the squares of the residuals in the regression limited to the sample period. The fit for this regression is exactly the same as the fit for the first T observations in the dummy variable regression, which means that the residuals are the same. And there are no residuals in the last P observations of the dummy variable regression because the inclusion of an observation-specific dummy in each observation guarantees a perfect fit in those observations. Hence RSS^D_{T+P} is exactly the same as RSS_T, and so for R^2.

(c) To evaluate quality of prediction one should compare predicted values of dependent variable according equation (2) and its actual values, the student writes them both in the table below

	2011	2012	2013	2014
H_t	6.4374	6.4697	6.4820	6.5073
$\hat{H}_t$	6.4539	6.4720	6.4846	6.5046

Using information from this table and equation (3) construct 95% confidence intervals for each year of the prediction period. How to use constructed confidence intervals for evaluation of the quality of prediction?

Solution. The predicted logarithm of expenditure on housing services in 2011 in the table was 6.4374. The standard error of the prediction error for that year was 0.0190. With 29 degrees of freedom, the critical value of t at the 5 percent significance level is 2.045, so we obtain the following 95 percent confidence interval for the prediction for that year: $6.4374 - 2.045 \times 0.0190 < y < 6.4374 + 2.045 \times 0.0190$, that is, $6.3985 < y < 6.4763$. We can see that the confidence interval does include the actual outcome, 6.4539, and thus, for that year at least, the prediction was satisfactory. The same is true for the remaining years in the prediction period.

(d) Finally the student estimates multiple regression model $H_t = \beta_1 + \beta_2 Y_t + \beta_3 P_t + u_t$ for the prediction period 2011-2014 ($RSS_P = 0.000218$). Now the student has information about three periods: the total period 1979-2014, the sample period 1979-2010, and the prediction period. Use conventional Chow test to compare the results of these three regressions. What is null hypothesis here in the context of the prediction? What are results of the test? What is your conclusion in the context of the prediction problem?

Running separate regressions for the two subperiods costs three degrees of freedom, and the number of degrees of freedom remaining after estimating six parameters (constant twice, coefficient of Y_t twice, and coefficient of P_t) is 30. Hence we obtain the following F statistic, which is distributed with 3 and 30 degrees of freedom:

$F(3,30)=\dfrac{(0.009218-[0.008958+0.000002])/3}{(0.008958+0.000002)/30}=0.29$ (also evidently insignificant).

Thus we conclude that there is no evidence of coefficient instability.

(e) To evaluate the quality of the prediction two different coefficients suggested by Theil could be used, $U_1=\dfrac{\sqrt{\frac{1}{n}\sum(H_{T+p}-\hat{H}_{T+p})^2}}{\sqrt{\frac{1}{n}\sum(H_{T+p})^2}+\sqrt{\frac{1}{n}\sum(\hat{H}_{T+p})^2}}$ and $U_2=\dfrac{\sqrt{\frac{1}{n}\sum(\Delta H_{T+p}-\Delta\hat{H}_{T+p})^2}}{\sqrt{\frac{1}{n}\sum(\Delta H_{T+p})^2}}$. Comment their properties in relation to the data under consideration.

Solution. Both coefficients are scale invariant. The Theil inequality coefficient U_1 always lies between zero and one, where zero indicates a perfect fit ($\hat{H}_{T+p}=H_{T+p}$).We have U = 1 (the "maximum inequality") if there is either a negative proportionality or if one of the variables is identically zero.

Some econometricians believe that U_1 has many disadvantages and prefer to use another coefficient $U_2=\sqrt{\dfrac{\frac{1}{h}\sum(\Delta\hat{y}_{T+p}-\Delta y_{T+p})^2}{\frac{1}{h}\sum(\Delta y_{T+p})^2}}$, based on evaluation of errors in differences and also suggested by Henry Theil.

The statistic U_2 has the advantage of possessing two natural calibration points. First it is equal to zero if the forecasts are perfectly accurate, and second, it is automatically equal to 1 for the naïve prediction of no change. If $\Delta\hat{H}_{T+p}=0$ for each forecast the numerator is the same as denominator. Since a forecasting model ought at a very minimum to be able to beat the forecast of no change, U_2 ought to lie between 0 and 1, its closeness to 0 being an indicator of its relative success.

In our case evaluation of U_1 gives $U_1=0.001334$, which corresponds to nearly perfect prediction.

The International College of Economics and Finance
Econometrics. Mid-year exam. 2017 October 26.

Part 2. (1 hour 30 minutes). Answer all questions (1,2,3) from section A and one (4 or 5) - from section B.

IMPORTANT: Start answering each question from the new page (ask for extra paper if necessary). Structure your answers in accordance with the structure of the questions. Testing hypotheses always state clearly null and alternative hypotheses provide critical value used for test, mentioning degrees of freedom and the significance level chosen for the test.

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

1. [15 marks]. Two students A and B are trying to answer the following question: what is better for future earnings – to study or to work, and so to get working experience. They collect data on 28 people working in different companies on their schooling S_i (in years), their working experience W_i (also in years) and current hourly earnings $EARN_i$ (in dollars). They also calculate the total number of years of active life spent on work or study $A_i = S_i + W_i$. It is assumed that one can not work and study at the same time.

The student A runs the following equation

$$\widehat{EARN}_i = -22.96 + 2.44 \cdot S_i + 0.92 \cdot W_i \qquad R^2 = 0.26 \qquad \textbf{(1)}$$
$$(17.48)\ (0.89)\qquad (0.85)$$

The student B using the same data estimates another equation

$$\widehat{EARN}_i = -22.96 + 2.44 \cdot A_i - 1.52 \cdot W_i \qquad R^2 = 0.26 \qquad \textbf{(2)}$$
$$(17.48)\ (0.89)\qquad (0.68)$$

(a) [5 marks]. Give interpretation to the coefficients of equation (1) and to all coefficients except that of W_i in the second equation. Explain why the coefficient of S_i in equation **(1)** is equal to the coefficient of A_i in equation **(2).** Explain why the constant terms in equations **(1)** and **(2)** are equal.

The interpretation of the first regression is straightforward: an extra year of study adds 2.44 dollars of hourly earnings, keeping working experience constant, and an extra year of working experience adds 0.92 dollars, keeping study years constant.

The interpretation of the coefficient of A_i shows the effect of one additional year of active life keeping working experience constant, so it is pure effect of one year of study as $A_i = S_i + W_i$: the change in A_i is fully achieved by the change of S_i. So the coefficients of S_i in the first equation should be the same as coefficient of A_i in the second equation.

The constant terms of both equations formally shows the position of the regression equations when both S_i and W_i are zero. So they should be the same (in fact the interpretation of the constant has no economic meaning as it could be supposed that schooling S_i cannot be zero.

(b) [5 marks]. The student A claims that spending one extra year to study is 2.5 times more useful for future earnings than spend it on work. Is he right? Give interpretation to the coefficient of W_i in the equation **(2)** and explain why it has negative sign. Is it possible to evaluate this coefficient using information from equation **(1)**?

Yes, the student is right: as S_i and W_i are both measured in years their coefficients are comparable, and marginal effect of study is $2.44/0.92 = 2.65$ times greater than the marginal effect of the same time period of working experience.

The coefficient of W_i in second equation indicates that an extra year of working experience *reduces* the earnings by 1.52 dollars per hour, which may at first sight seem nonsensical. However, the interpretation is conditional on holding A_i constant, and an increase in W_i, holding A_i constant, means a reduction in S_i. So the coefficient is the net value of transferring the year from study to working experience, and since the first equation indicates that working experience is less valuable than study, the net effect is negative. It is the difference between the coefficients in the first equation: 2.44-0.92=1.52.

We can see this algebraically by writing the first equation as

$$EARN = \beta_1 + \beta_2 S + \beta_3 W + u$$

and then reparameterizing it as

$$\begin{aligned} EARN &= \beta_1 + \beta_2 S - \beta_2 W + \beta_2 W + \beta_3 W + u \\ &= \beta_1 + \beta_2 A + (\beta_2 - \beta_3)W + u \end{aligned}$$

(c) [5 marks]. Evaluate the significance of the coefficients. Evaluate the significance of the equation **(1)** using F-statistic**.** Is it possible to get F-statistic for equation **(2)** without calculation? Why the equations **(1)** and **(2)** have the same R^2? Why the standard error of the coefficient of S_i in the equation **(1)** is equal to the standard error of the coefficient of A_i in the equation **(2)**?

The number of degrees of freedom for both equations is the same $28 - 3 = 25$.

$$EARN = \beta_1 + \beta_2 S + \beta_3 W + u$$

Hypotheses: (for example for S)

$H_0 : \beta_2 = 0$

$H_a : \beta_2 \neq 0$

t-statistics for coefficients are

for equation (1): $t_S = 2.44/0.89 = 2.74$ and $t_W = 0.92/0.85 = 1.08$,

for equation (2): $t_A = 2.44/0.89 = 2.74$ and $t_W = -1.52/0.68 = 2.24$.

The critical values are $t_{crit}(5\%, df = 25) = 2.060$ and $t_{crit}(1\%, df = 25) = 2.787$.

So the coefficient of S_i in equation (1) and the coefficient of A_i in equation (2) are significant at 5%, while W_i is insignificant in (1) and significant at 1% in (2).

The equation (1) is significant at 5% significance level as $F = \dfrac{0.26/2}{(1-0.26)/25} = 4.39$ while $F_{crit}(5\%, 2, 25) = 3.39$ and $F_{crit}(1\%, 2, 25) = 5.61$.

Both equation (1) and (2) have the same R^2 because they are evaluated on the base of the same data on the variables $EARN_i$, S_i and W_i. (This clearly follows from reparametrization in (b)).

The standard error of the coefficient of S_i in the equation **(1)** is equal to the standard error of the coefficient of A_i in the equation **(2)** because the coefficient is in fact the same as it has been shown in (b), and it is estimated using the same data with the same dependent and the same explanatory variables.

2. [15 marks]. Consider the following regression model

$$Y_i = \beta_1 + \beta_2 X_i + u_i, \quad i = 1, \ldots, n, \qquad \textbf{(1)}$$

where the values of X_i are assumed to be non-stochastic and u_i satisfy Gauss-Markov conditions.

(a) [6 marks]. The regression **(1)** is estimated using OLS, and the following characteristics of regression are obtained: TSS, RSS, R^2, F-statistic and t-statistic for the coefficient β_2. If we add additional observation to the sample, how does this affect the values of TSS, RSS, R^2, F-statistic and t-statistic for the coefficient β_2? Brief explanations are required for each point.

We accept without proof that TSS is generally speaking is increases as new observation is included in the sample (the proof is at the end of this section).

RSS also increases, the proof is simple. Let for n observations residual sum of squares is equal to RSS_n^*. According to the principle of the least squares RSS_n^* is minimal for the sample points $(X_i, Y_i), i = 1, \ldots, n$. If new observation (X_{n+1}, Y_{n+1}) is added then the new value of $RSS_{n+1} = \sum_1^{n+1}(Y_i - \hat{Y}_i)^2$ could be decomposed into two parts: the sum of squared deviations $RSS_n = \sum_1^n (Y_i - \hat{Y}_i)^2$ for the 'old' sample $(X_i, Y_i), i = 1, \ldots, n$ and square of the deviation $(Y_{n+1} - \hat{Y}_{n+1})^2$ for the 'new' observation (X_{n+1}, Y_{n+1}). So

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

As $R^2 = 1 - \dfrac{RSS}{TSS}$ and both RSS and TSS are increasing with the inclusion of the new observation, the result is unpredictable. From the practical side it can be clear explained: if new observation (X_{n+1}, Y_{n+1}) lies on the regression line obtained using the sample $(X_i, Y_i), i = 1, \ldots, n$ (or is close to it), R^2 increases as in this case correlation becomes greater (in absolute value). If the new observation is outlier and lies far from the former regression line the general fit becomes worse and so R^2 decreases.

The same conclusion can be made about the value of F-statistic, as it can be directly derived from R^2: $F = \dfrac{R^2}{1 - R^2}(n-2)$, and also about the value of t-statistic as for the simple regression $t^2 = F$.

The most difficult question is about TSS. By definition $TSS_n = \sum_1^n (Y_i - \bar{Y}_n)^2$ where $\bar{Y}_n = \dfrac{\sum_1^n Y_i}{n}$. We show that $TSS_{n+1} \geq TSS_n$. The proof is based on the well known property of the mean: if C is any number then $\sum_1^n (Y_i - C)^2 = \sum_1^n (Y_i - \bar{Y})^2 + n(C - \bar{Y})^2$.

Proof: $\sum_1^n (Y_i - C)^2 - \sum_1^n (Y_i - \bar{Y})^2 = \sum_1^n ((Y_i - C)^2 - (Y_i - \bar{Y})^2) = \sum_1^n (Y_i^2 - 2CY_i + C^2 - Y_i^2 + 2Y_i\bar{Y} - \bar{Y}^2) =$
$= -2Cn\bar{Y} + nC^2 + 2n\bar{Y}^2 - n\bar{Y}^2 = n(\bar{Y}^2 - 2C\bar{Y} + C^2) = n(C - \bar{Y})^2$.

Now let $C = \bar{Y}_{n+1}$, then

$$TSS_{n+1} = \sum_1^{n+1}(Y_i - \bar{Y}_{n+1})^2 = \sum_1^{n}(Y_i - \bar{Y}_{n+1})^2 + (Y_{n+1} - \bar{Y}_{n+1})^2 =$$

$$= \sum_1^{n}(Y_i - \bar{Y}_n)^2 + n(\bar{Y}_{n+1} - \bar{Y}_n)^2 + (Y_{n+1} - \bar{Y}_{n+1})^2 \geq \sum_1^{n}(Y_i - \bar{Y}_n)^2 = TSS_n$$

Comment: Inequality $\sum_1^n (Y_i - C)^2 \ge \sum_1^n (Y_i - \overline{Y})^2$. (where C is any number) can be also proved on the base of Least Squares Principle (as it has been done in (c)). The remainder of the proof is the same.

(b) [4 marks]. Now additional explanatory variable is added to regression **(1)** to get regression **(2)**:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 Z_i + u_i, \quad i = 1, \ldots, n, \qquad \textbf{(2)}$$

The regressions **(1)** and **(2)** are estimated using OLS on the data of the sample of initial size $i = 1, \ldots, n$, and the following characteristics of regressions are obtained: for regression **(1):** TSS_1, RSS_1, R_1^2, $\overline{R}_{1\,adj}^2$, F_1 - F-statistic for the whole equation; for regression **(2):** TSS_2, RSS_2, R_2^2, $\overline{R}_{2,\,adj}^2$, F_2 - F-statistic for the whole equation. What can be said about the relationship between the corresponding characteristics of these two regressions **(1)** and **(2)**? Give explanations for each pair of characteristics compared.

As both equations have the same dependent variable $TSS_1 = TSS_2$.

Equation (1) can be considered as the restricted version of the equation (2) with the restriction $\beta_3 = 0$, and so $RSS_1 \ge RSS_2$.

As it is well known R^2 always increases with the adding of new variable, so $R_1^2 \le R_2^2$.

$\overline{R}_{adj}^2$ is defined as $R^2 - \frac{k-1}{n-k}(1-R^2)$ (or $1 - \frac{n-1}{n-k}(1-R^2)$) where n is the number of observations and k is the number of estimated parameters. As in this formula along with increasing R^2 there is also a penalty in the form $-\frac{k-1}{n-k}(1-R^2)$ the result is unpredictable. It is known that $\overline{R}_{adj}^2$ increases only if t-statistic of added variable is greater than unity.

As for F-statistic defined as $F = \frac{R^2/(k-1)}{(1-R^2)}(n-k)$ two factors act in the opposite direction: with adding new varianle R^2 influence positively, while increased number of estimated parameters (in $(k-1)$ and in $(n-k)$) influence negatively, so the result is unpredictable.

(c) [5 marks]. Now the whole sample is divided into two subsamples **(A)** $i = 1, \ldots, m$, and **(B)** $i = m+1, \ldots, n$. The regression **(2)** is estimated using OLS for these subsamples: TSS_A, RSS_A, R_A^2 - for subsample **(A)**, and TSS_B, RSS_B, R_B^2 - for subsample **(B).** What are relations between TSS_2, RSS_2, R_2^2 for the whole sample and corresponding sums $TSS_A + TSS_B$, $RSS_A + RSS_B$, and $R_A^2 + R_B^2$. Explanations are expected in each case.

Again the most difficult question is about TSS. By definition $TSS_n = \sum_1^n (Y_i - \overline{Y}_n)^2$ where $\overline{Y}_n = \frac{\sum_1^n Y_i}{n}$.

Solving the problem $\sum_1^n (Y_i - A)^2 \to \min_A$ (using derivatives) it can be shown that $A = \overline{Y}_n$, so the position of $\overline{Y}_n$ provides minimum value to the sum of squares. Let $\overline{Y}_A$ and $\overline{Y}_B$ are means for two subsamples A and B: $\overline{Y}_A = \dfrac{\sum_1^m Y_i}{m}$ and $\overline{Y}_B = \dfrac{\sum_{m+1}^n Y_i}{n-m}$. Dividing total sum of squares into two parts corresponding to the two subsamples A and B we get from here

$$TSS_2 = \sum_1^n (Y_i - \overline{Y}_n)^2 = \sum_1^m (Y_i - \overline{Y}_n)^2 + \sum_1^n (Y_i - \overline{Y}_n)^2 \geq \sum_1^m (Y_i - \overline{Y}_A)^2 + \sum_1^n (Y_i - \overline{Y}_B)^2 = TSS_A + TSS_B \,.$$

As for RSS we also use the decomposition of the total sum into two parts corresponding to each subsample

$$RSS_2 = \sum_1^n (Y_i - \hat{Y}_n)^2 = \sum_1^m (Y_i - \hat{Y}_n)^2 + \sum_1^n (Y_i - \hat{Y}_n)^2 \geq \sum_1^m (Y_i - \hat{Y}_A)^2 + \sum_1^n (Y_i - \hat{Y}_B)^2 = RSS_A + RSS_B \,.$$

As $R^2 = 1 - \dfrac{RSS}{TSS}$ and both RSS and TSS are decreasing with the dividing into two subsamples the result for R^2 is unpredictable.

3. [15 marks]. The rise in prices for public transport leads to lower corporate earnings, as people tend to choose cheaper alternatives. The student tries to find the best form of dependence of the volume of transportation TR_i (in millions of dollars) for 50 transportation companies from the prices of transportation P_i (in cents per one kilometer of transportation). She runs regressions (1-4) (linear, logarithmic and semi-logarithmic functions), she also runs two auxiliary regressions (5-6) performing Zarembka transformation (variable TRZ_i is defined as $TRZ_i = TR_i / \sqrt[n]{TR_1 \cdot TR_2 \cdot \ldots \cdot TR_n}$):

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Dependent variable	TR_i	TR_i	$\log(TR_i)$	$\log(TR_i)$	TRZ_i	TRZ_i
Independent variable\Constant	8.74	12.26	2.175	2.635	1.171	1.641
P_i	-0.339	-	-0.0045		-0.0045	
$\log(P_i)$	-	-1.362	-	-0.179	-	-0.179
R^2	0.638	0.738	0.665	0.755	0.638	0.738
RSS	4.481	3.247	0.068	0.051	0.080	0.058

(a) [5 marks]. Explain the differences in the values of the slope coefficients in regression (1) and (4) giving interpretation to both regressions. Explain using some math why your interpretation of regression (4) is correct.

The equations (1) and (4) have different specifications: (1) is linear while (4) is double-logarithmic, so the interpretation of their slope coefficients are different

For (1) coefficient of P_i equal to -0.339 shows the marginal effect of prices on the volume of transportation, namely if the prices rise by 1 cent per one kilometer the volume of transportation drops by 339 thousands of dollars.

For (4) coefficient of $\log(P_i)$ equal to -0.179 shows the price elasticity of the volume of transportation, namely if the price per one kilometer of transportation rises by 1 percent the volume of transportation drops by 0.179 percent. They are different as they have different economic meaning, but they both are negative expressing negative influence of the rising of prices on the volume of transportation.

(explanations) It is sufficient here to present one of the following two ways of mathematical reasoning.

For model (4). Direct method.

Let the dependence of volume of transportation of prices has specification

$Y = \beta_1 X^{\beta_2}$. Then

$$\frac{dY}{dX} = \beta_1 \beta_2 X^{\beta_2 - 1}$$

On the other hand

$$\frac{Y}{X} = \frac{\beta_1 X^{\beta_2}}{X} = \beta_1 X^{\beta_2 - 1}$$

So elasticity is $\dfrac{dY/dX}{Y/X} = \dfrac{\beta_1 \beta_2 X^{\beta_2 - 1}}{\beta_1 X^{\beta_2 - 1}} = \beta_2$

OR

For model (4). Method of differentials.

Taking logarithms of both sides of $Y = \beta_1 X^{\beta_2}$ we get the function of the form $\log Y = a + b \cdot \log X$ then

$$\frac{dY}{Y} = d(a + b \cdot \log X) = b \cdot \frac{dX}{X}$$, from here

$$b = \frac{\frac{dY}{Y}}{\frac{dX}{X}} = \frac{dY}{dX} \cdot \frac{X}{Y} = \frac{\frac{dY}{Y} \cdot 100\%}{\frac{dX}{X} \cdot 100\%}$$

setting $\frac{dX}{X} \cdot 100\% = 1\%$ obtain $b = \frac{dY}{Y} \cdot 100\%$

so b shows the percentage increase of Y when X increases by one percent.

(b) [5 marks]. Explain the differences in the values of the slope coefficients in regression (2) and (3) giving interpretation to both regressions. Explain using some math why your interpretation is correct (for one of the regressions 2-3).

The equations (2) and (3) are both semi-logarithmic regressions but of different specifications:
Equation (2) is linear-logarithmic regression while (3) is log-linear regression, so the interpretation of their slope coefficients are different.
For (2) coefficient of $\log(P_i)$ equal to -1.362 shows that if the prices rise by 1% the volume of transportation drops by $\frac{-1.362}{100} = -0.01362$ millions of dollars or by 13.62 thousands of dollars on average.
For (3) coefficient of P_i equal to -0.0045 shows that if the price of transportation rises by 1 cent per one kilometer the volume of transportation drops by $-0.0045 \cdot 100 = -0.45$ percent on average.

For model (2). Method of differentials.
Let $Y = a + b \cdot \log X$ then

$$dY = d(a + b \cdot \log X) = b \cdot \frac{dX}{X},$$ from here

$$b = \frac{dY}{\frac{dX}{X}} \quad \text{or} \quad \frac{b}{100} = \frac{dY}{\frac{dX}{X} * 100(\%)}$$

setting $\frac{dX}{X} \cdot 100\% = 1\%$ obtain $\frac{b}{100} = dY$

If you divide b by 100, the resulting number will show the increase of Y, provided X will increase by one percent (the units of measurement should always be clearly indicated)

For model (3). Method of differentials.
Let $\log Y = a + bX$ then

$$\frac{dY}{Y} = d(a + bX) = bdX,$$ from here

$$\frac{dY}{Y} \cdot 100\% = b \cdot 100\% dX \quad \text{or} \quad b \cdot 100\% = \frac{\frac{dY}{Y} \cdot 100\%}{dX}$$

setting $dX = 1$ obtain $b \cdot 100\% = \frac{dY}{Y} \cdot 100\%$

If you multiply b by 100, the resulting number will show the percentage increase of Y, provided X will increase by one unit (the units of measurement should always be clearly indicated)

Other ways of explaining how to interpret the coefficients are also welcomed. (the students who did this hey were awarded by bonus marks).
For example for model (2). Direct method.
Assume $X = 100$ then
$Y(100) = a + b \cdot \log 100$
Now $X = 101$, then
$Y(101) = a + b \cdot \log 101$
Subtracting first equality from the second
$\Delta Y = Y(101) - Y(100) = b \cdot \log 101 - b \cdot \log 100 = b \cdot \log \frac{101}{100} = b \cdot \log 1.01 \approx b \cdot 0.01 = \frac{b}{100} \cdot 1\%$ so
The coefficient b divided by 100 shows the increase of Y (in absolute units) in response to the 1% increase of X .

(c) [5 marks]. Which pairs of regression are comparable directly (without Zarembka transformation). Compare the regressions in such pair using appropriate data. Which regressions becomes comparable after Zarembka transformation? Compare any two of these regressions using appropriate test.

(c4) The regressions (1) and (2) and regressions (3) and (4) are comparable directly, as in these pairs the dependent variables are the same and so their TSS's (Total Sums of Squares) are the same. So the comparison can be made on the basis of their RSS's or determination coefficients: (2) is better than (1) and (3) is better than (4) (we do not discuss here the significance of the differences).
To compare (1) and (3) and to compare (2) and (4), and also to estimate the significance of the difference in their quality one need to perform Box-Cox test on the base of the auxiliary regressions (5) and (6) correspondingly that uses Zarembka transformation of the dependent variable.
As it is known the value $\chi^2 = (\frac{n}{2})\left|\log\left(\frac{RSS1}{RSS2}\right)\right|$ has χ^2-distibution with 1 degrees of freedom.
For the comparison of (1) and (3) (and so using (3) and (5)) we get
$\chi^2 = (\frac{50}{2})\left|\log\left(\frac{0.080}{0.68}\right)\right| = 4.063$ what is greater than 5% critical value of chi-squared for 1 degree of freedom 3.84. So the diffefence in quality of regressions is significant and the student shoud choose the regression (3) and drop out the regression (1) as having less quality.
The same test in pair (2) and (4) (using for the test (4) and (6)) gives
$\chi^2 = (\frac{50}{2})\left|\log\left(\frac{0.058}{0.051}\right)\right| = 3.215$ what is less than 5% critical value of chi-squared for 1 degree of freedom 3.84. So the diffefence in quality of regressions is insignificant and the student can choose any of the regressions (2) or (4). For example, she could choose (4) as it has clear economic interpretation using elasticity and also has less RSS.

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

4. [30 marks]. Consider the model $Y_i = \beta_1 + \beta_2 X_i + u_i$, $i = 1, \dots, n$, where $E(u_i) = 0$, $E(u_i^2) = \sigma_u^2 = const.$; $E(u_i u_j) = 0$, $i \neq j$; The values of $X_i, i = 1, \dots, n$ are assumed to be non-stochastic with non-zero sample variance $\mathrm{Var}(X)$.

(a) [4 marks]. Show that $b_1 = \beta_1 + \sum_{i=1}^{n} g_i u_i$, where $g_i = \frac{1}{n} - a_i \overline{X}$ where $a_i = \dfrac{X_i - \overline{X}}{\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2}$.

$$b_1 = \overline{Y} - b_2\overline{X} = (\beta_1 + \beta_2\overline{X} + \overline{u}) - \overline{X}(\beta_2 + \sum a_i u_i) =$$
$$= \beta_1 + \frac{1}{n}\sum u_i - \overline{X}\sum a_i u_i = \beta_1 + \sum(\frac{1}{n} - a_i\overline{X})u_i = \beta_1 + \sum g_i u_i$$, where $g_1 = \frac{1}{n} - a_i\overline{X}$

(b) [4 marks]. Prove that $\sum_{i=1}^{n} g_i = 1$, and $\sum_{i=1}^{n} g_i X_i = 0$.

$$\sum g_i = \sum(\frac{1}{n} - a_i\overline{X}) = \sum(\frac{1}{n}) - \overline{X}\sum a_i = 1 - \overline{X}\cdot 0 = 1 \quad (\text{as } \sum a_i = 0)$$
$$\sum g_i X_i = \sum(\frac{1}{n} - a_i\overline{X})X_i = \sum(\frac{X_i}{n}) - \overline{X}\sum a_i X_i = \overline{X} - \overline{X} = 0 \quad (\text{as } \sum a_i X_i = 1)$$

(c) [7 marks]. Show that OLS estimator of the intercept b_1 is linear unbiased estimator of β_1

Part one: b_1 is unbiased.

Using the decomposition $b_1 = \beta_1 + \sum_{i=1}^{n} g_i u_i$ where we get

$$E(b_1) = E\left(\beta_1 + \sum c_i u_i\right) = \beta_1 + \sum c_i E(u_i) = \beta_1$$

Part two: b_1 is linear function of Y_i.

We know that $b_2 = \sum_{i=1}^{n} a_i Y_i$, where $a_i = \dfrac{(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$ so.

$$b_1 = \overline{Y} - b_2\overline{X} = \frac{1}{n}\sum_{i=1}^{n} Y_i - \overline{X}\sum_{i=1}^{n} a_i Y_i = \sum_{i=1}^{n}(\frac{1}{n} - a_i\overline{X})Y_i = \sum_{i=1}^{n} g_i Y_i$$

where g_i are defined as above $g_1 = \frac{1}{n} - a_i\overline{X}$.

(d) [7 marks]. Prove that the variance of OLS estimator of the intercept is

$$\sigma_{b_1}^2 = \sigma_u^2 \sum g_i^2 = \sigma_u^2\left\{\frac{1}{n} + \frac{\overline{X}^2}{\sum\left(X_i - \overline{X}\right)^2}\right\}$$

$b_1 = \beta_1 + \sum g_i u_i$, where $g_1 = \frac{1}{n} - a_i\overline{X}$. Hence

$$\sigma_{b_1}^2 = E\left[\left(\sum c_i u_i\right)^2\right] = \sigma_u^2 \sum c_i^2 = \sigma_u^2\left(n\frac{1}{n^2} - 2\frac{\overline{X}}{n}\sum a_i + \overline{X}^2 \sum a_i^2\right).$$

As it is known, $\sum a_i = 0$ and $\sum a_i^2 = \frac{1}{\sum\left(X_i - \overline{X}\right)^2}$.

Hence $\sigma_{b_1}^2 = \sigma_u^2\left\{\frac{1}{n} + \frac{\overline{X}^2}{\sum\left(X_i - \overline{X}\right)^2}\right\}$.

The alternative way to prove this is to use directly properties of variance
It should be mentioned that if u_i are uncorrelated then Y_i are also uncorrelated

Using the decomposition $b_1 = \beta_1 + \sum_{i=1}^{n} g_i u_i$ where $g_1 = \frac{1}{n} - a_i \overline{X}$ and $a_i = \frac{X_i - \overline{X}}{\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2}$ we get

$$\operatorname{var}(b_1) = \operatorname{var}(\beta_1 + \sum_{i=1}^{n}(\frac{1}{n} - a_i\overline{X})u_i) = \operatorname{var}(\sum_{i=1}^{n}(\frac{1}{n} - a_i\overline{X})u_i) = \sum_{i=1}^{n}(\frac{1}{n} - a_i\overline{X})^2 \operatorname{var}(u_i)$$

(as u_i are uncorrelated according to assumptions of model A or G-M Conditions)

$$= \sigma_u^2\sum_{i=1}^{n}(\frac{1}{n} - a_i\overline{X})^2 = \sigma_u^2\sum_{i=1}^{n}(\frac{1}{n} - a_i\overline{X})^2 = \sigma_u^2\sum_{i=1}^{n}(\frac{1}{n^2} - 2\frac{a_i\overline{X}}{n} + a_i^2(\overline{X})^2)$$

$$= \sigma_u^2(\frac{n}{n^2} - 2\frac{\overline{X}\sum_{i=1}^{n} a_i}{n} + (\overline{X})^2\sum_{i=1}^{n} a_i^2) = \sigma_u^2(\frac{1}{n} + \frac{(\overline{X})^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}) \text{ as } \sum_{i=1}^{n} a_i = 0 \text{ and } \sum_{i=1}^{n} a_i^2 = \frac{1}{\sum_{i=1}^{n}(X_i - \overline{X})^2}.$$

(e) [8 marks]. Prove that $\operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) = \frac{-\overline{X}\sigma_u^2}{\sum(X_i - \overline{X})^2}$.

$$\operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) = \operatorname{cov}(\overline{Y} - \hat{\beta}_2\overline{X}, \hat{\beta}_2) = \operatorname{cov}(\overline{Y}, \hat{\beta}_2) - \overline{X}\operatorname{cov}(\hat{\beta}_2, \hat{\beta}_2) = 0 - \overline{X}\operatorname{var}(\hat{\beta}_2) = \frac{-\overline{X}\sigma^2}{\sum(X_i - \overline{X})^2}.$$

The equality $\operatorname{cov}(\overline{Y}, \hat{\beta}_2) = 0$ can be proven in the following way

$$\operatorname{cov}(\overline{Y}, \hat{\beta}_2) = \operatorname{cov}\left(\sum\frac{1}{n}Y_i, \sum a_i Y_i\right) = \sum\left(\frac{a_i}{n}\right)\operatorname{var}(Y_i) + \sum_{i<j}\left(\frac{a_j}{n}\right)\operatorname{cov}(Y_i, Y_j) = \frac{\sigma_u^2}{n}\sum a_i = 0$$

as $\operatorname{cov}(Y_i, Y_j) = 0$ where $i \neq j$.

We use here the decomposition $\hat{\beta}_2 = \sum a_i Y_i$, where $a_i = \frac{(X_i - \overline{X})}{\sum(X_i - \overline{X})^2}$, and $\sum a_i = 0$.

5. [30 marks]. A student of ICEF preparing her diploma paper studies the dependence of the prices on the paintings P_i (in thousands of dollars) on various factors, in particular on the "age" of the painting AGE_i (in decades from the current year to the year of creation of the piece of art) and on the size of the canvas S_i (in square feets). She collected data on 19 paintings sold at a certain auction to estimate equation **(1)**

$$\hat{P}_i = \underset{(0.57)}{2.35} + \underset{(0.011)}{0.028} \cdot AGE_i + \underset{(0.020)}{0.037} \cdot S_i \qquad R^2 = 0.325,\quad RSS = 0.417 \qquad \textbf{(1)}$$

Correlation between variables AGE_i and S_i was -0.93.

(a) [5 marks]. Are two variables AGE_i and S_i significant taken by one and together, taking into account that the student was not sure in the signs of coefficients before the regression was estimated. Give the interpretation to all coefficients of this equation.

Theoretical equation

$$P_i = \beta_1 + \beta_2 \cdot AGE_i + \beta_3 \cdot S_i + u_i$$

To test the significance of the slope coefficients $\begin{pmatrix} H_0 : \beta_i = 0, \\ H_a : \beta_i \neq 0 \end{pmatrix}$ $i = 2, 3$ let us evaluate t-statistics for the coefficients: $t_{AGE} = \frac{0.028}{0.011} = 2.545$, $t_S = \frac{0.037}{0.02} = 1.85$, $t(crit.,5\%, 2sided, df = 16) = 2.12$, so AGE_i is significant while S_i is not.

To test joint significance of two variables AGE_i and S_i taken together first write the theoretical equation

$$P_i = \beta_1 + \beta_2 \cdot AGE_i + \beta_3 \cdot S_i + u_i$$

Null hypothesis is $H_o : \beta_2 = 0, \beta_3 = 0$

$F = \frac{R^2/2}{(1-R^2)/df} = \frac{0.325/2}{(1-0.325)/16} = 3.852$ while $F(crit.,5\%, df = 2,16) = 3.63$, so equation as a whole is significant.

The coefficient of AGE shows the marginal effect of distance when S_i is fixed, while coefficient of S_i shows the marginal effect of canvases size on the price when the age of the artwork is fixed. This means for example that the price rises by 28 dollars on average if the age rises by decade keeping size of the canvas fixed.

Similarly for the coefficient of the size of painting. The intercept formally has no interpretation as there can be no painting of zero size.

(b) [8 marks]. Since the variable S_i was insignificant the student decided to exclude it from equation.

$$\hat{P}_i = \underset{(0.076)}{3.36} + \underset{(0.0046)}{0.0091} \cdot AGE_i \qquad R^2 = 0.1875,\quad RSS = 0.502 \qquad \textbf{(2)}$$

Use method of confidence intervals to test the significance of the regression coefficients. Help the student to get full interpretation of the obtained equation, and explain based on this interpretation why coefficient of variable AGE_i is now smaller than in equation **(1)**, and why the standard error of the coefficient of the variable AGE_i in **(2)** is smaller than in **(1)**?.

Confidence interval $b_i \pm t(crit.,95\%,df=17)\cdot s.e.(b_1)$ is here $-0.0091 \pm 2.110 \cdot 0.0046$ or $(-0.0188;\ 0.0006)$. As confidence interval contains zero, the slope coefficient equal to zero is one of the acceptable hypotheses. Coefficient is insignificant.

99% confidence interval for intercept is $3.16 = 3.36 - 2.898 \cdot 0.07;\ 3.36 + 2.898 \cdot 0.07 = 3.56$ $(3.16;\ 3.56)$ so the intercept is significant..

Coefficient 0.0091 of AGE_i shows marginal effect of the age: if age of the painting increases by 1 decade the prices increase by 9.1 dollars on average..

The estimate of the price of the contemporary piece of art is 3.36 thousands of dollars..

This interpretation of the intercept is valid if the sample is really contains data on contemporary paintings. It should be noted that since the constant term of equation is significant, and the coefficient is not significant, the equation can be interpreted also as an estimate of the average price sold at the auction..

The coefficient of the AGE_i is now smaller than in (1), that can be explained based on the interpretation. It is said that the correlation between variables AGE_i and S_i is -0.93, so variables AGE_i and S_i in the sample tend to move in the opposite direction affecting prices of paintings in the opposite way. The result of a compromise between the opposing influences is the slope coefficient in equation **(2)** which is less than in **(1)** where pure effect of age is considered.

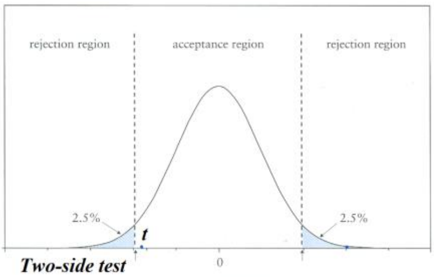
The theoretical value for the standard error of the coefficient of AGE_i in **(2)** is $\text{s.e.}(b_2) = \sqrt{\dfrac{s_u^2}{\sum\left(X_{2i} - \bar{X}_2\right)^2}}$

while the one for equation **(1)** is $\text{s.e.}(b_2) = \sqrt{\dfrac{s_u^2}{\sum\left(X_{2i} - \bar{X}_2\right)^2} \times \dfrac{1}{1 - r^2_{X_2,X_3}}}$. As $r^2_{X_2,X_3} = (-0.93)^2$ the factor is $\text{s.e.}(b_2) = \sqrt{\dfrac{1}{1 - r^2_{X_2,X_3}}} = \sqrt{\dfrac{1}{1 - 0.93^2}} = 2.72$ when actual ratio is $\dfrac{0.011}{0.0046} = 2.39$.

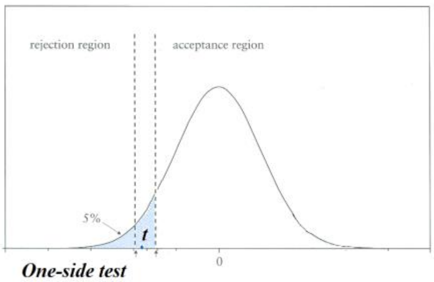
The observed situation when equation as a whole is significant while one of the variables (S_i) is insignificant indicates on the presence of multicollinearity caused probably by the high correlation between AGE_i and S_i equal to -0.93.

(c) [5 marks] Explain the researcher why having well-founded assumptions about the signs of regression coefficients, one can in some cases justify their significance. Does it really help here in estimating equations **(1)** and **(2)**? Explain how to use this method correctly stating clearly the null and alternative hypotheses.

Using one-sided tests $\begin{pmatrix} H_0 : \beta = 0, \\ H_a : \beta < 0 \end{pmatrix}$ or $\begin{pmatrix} H_0 : \beta = 0, \\ H_a : \beta > 0 \end{pmatrix}$ could help in some situations to make significant the coefficient that is insignificant when using two-sided test. For example in the situation $\begin{pmatrix} H_0 : \beta = 0, \\ H_a : \beta < 0 \end{pmatrix}$ if t-statistics is negative but is in acceptance region for 5% two-sided test it could be in rejection region for one-sided test as probability to the left of the critical value is doubled now so the critical value is shifted to the right (see picture).



Two-side test



One-side test

For equation **(2)** $t^{crit}_{5\%}(\text{one}-\text{sided},17)=1.740$, so if pair of hypotheses $\begin{pmatrix} H_0:\beta=0, \\ H_a:\beta>0 \end{pmatrix}$ is chosen, the coefficient of AGE_i becomes significant at 5% level. ($t_{AGE}=\dfrac{0.0091}{0.0046}=1.978$)

For equation **(1)** $t^{crit}_{5\%}(\text{one}-\text{sided},16)=1.746$, so coefficient of S_i becomes significant 5% level ($t_S=\dfrac{0.037}{0.02}=1.85$).

(d) [7 marks]. The supervisor advised the student to take into account also the availability of the provenance (a history of possession of an artwork, confirming the authenticity of the object): the variable $PV_i=1$ if the provenance is present, and $PV_i=0$ otherwise. The student estimates two additional equations **(3), (4)**

$$\hat{P}_i=\underset{(0.71)}{0.87}+\underset{(0.016)}{0.062}\cdot AGE_i+\underset{(0.024)}{0.082}\cdot S+\underset{(0.13)}{0.36}\cdot PV_i \qquad R^2=0.554,\ RSS=0.276 \qquad \textbf{(3)}$$

and

$$\hat{P}_i=\underset{(0.75)}{-0.097}+\underset{(0.017)}{0.085}AGE_i+\underset{(0.025)}{0.11}\cdot S+\underset{(3.55)}{2.69}PV_i-\underset{(0.05)}{0.07}PV_i\cdot AGE_i-\underset{(0.14)}{0.08}PV_i\cdot S_i \qquad R^2=0.691,\ RSS=0.191 \qquad \textbf{(4)}$$

What is the difference in interpretation of coefficients of the equations **(1), (3), (4)**? Is the availability of the provenance significant according **(3)** and **(4)**?

In equation **(3)** the constant term 0.87 relates to the painting without provenance, while for painting with provenance it is equal to $0.87+0.36=1.23$. The marginal effects of AGE_i and S_i are supposed to be the same for painting with probenance and without it.
In equation **(4)** the constant term -0.097 also relates to the painting without provenance, while for painting with provenance it is equal to $-0.097+2.69=2.593$. The marginal effects of AGE_i and S_i are now depend on availability of provenance: for painting without provenance the marginal effects of AGE_i and S_i are correspondingly 0.085 and 0.11, while for painting with provenance they are $0.085-0.07=0.015$ and $0.11-0.08=0.03$.

In equation **(3)** PV_i is significant at 5% according to t-test ($t=\dfrac{0.36}{0.13}=2.77$). To test the significance of provenance one have to use F-test for the group of variables PV_i, $PV_i\cdot AGE_i$, and $PV_i\cdot S_i$:

$$F=\frac{(R^2_U-R^2_R)/3}{(1-R^2_U)/df_U}=\frac{(0.691-0.325)/3}{(1-0.691)/13}=5.13 \ \left(\text{or } F=\frac{(RSS_R-RSS_U)/3}{RSS_U)/df_U}=\frac{(0.417-0.191)/3}{0.191/13}=5.13\right)$$

while $F(crit.,5\%,df=3,13)=3.41$, so the provenance is significant.

(e) [5 marks]. The alternative approach to the detection of the significance of the provenance is Chow test. Describe in details this approach. Suppose that equation **(1)** estimated for paintings without provenance gives the value of RSS equal to 0.191, while for those with provenance the value of RSS is 0.000122. Do Chow test and compare the results with the tests above in (d).

To perform Chow test one need to run the equation

$$P_i=\beta_1+\beta_2\cdot AGE_i+\beta_3\cdot S_i+u_i$$

for the whole sample (equation (2) with RSS equal to 0.417, and then to run the same equation for two subsamples corresponding to $PV_i = 0$ (with RSS = 0.191) and to $PV_i = 1$ (with RSS = 0.000122).

The Chow test gives the result

$F = \dfrac{(RSS_{TOT} - (RSS_{PV=0} + RSS_{PV=1})/3}{(RSS_{PV=0} + RSS_{PV=1})/(19-2*3)} = \dfrac{(0.417-(0.191-0.000122))/3}{(0.191-0.000122)/13} = 5.13$, which is equivalent to F-test in **(d).**

The International College of Economics and Finance
Econometrics – 2012-2013.
Mid-term exam 2013. March 29.

SUGGESTED SOLUTIONS AND MARKING

General instructions. **Answer all questions of the problem 1 and any three of the problems 2-6. The weight of the obligatory problem 1 is 25% of the exam; three other questions add 25% each. You are advised to divide your time accordingly. Structure your answers in accordance with the structure of the questions. When testing hypotheses always state clearly null and alternative hypotheses provide critical value used for test, mentioning degrees of freedom and the significance level chosen for the test.**

Part 2. (2 hours 15 minutes). Answer any three of the five problems (2-6).

1. Data on household expenditure on clothing (*CLOTH*) in dollars is obtained from a survey of 400 households in a two week period. In addition data on household income (*INC*) in dollars, the location of households (REGION) and the size of the family (*SIZE*) in number of persons was collected. An econometrician wants to test the hypothesis that the effects of income and family size are different in the North of the country as compared to the South. From a variable which denotes the region of residence of the household the data is divided into two sets, one for the North (200 households) and one for the South (200 households). The equations to be estimated are:

	REGION
$CLOTH_i = \alpha_0 + \alpha_1 INC_i + \alpha_2 SIZE_i + u_{1i}$	(North)
$CLOTH_i = \beta_0 + \beta_1 INC_i + \beta_2 SIZE_i + u_{2i}$	(South)

Where u_{1i} and u_{2i} are unobserved disturbances.

(a) Describe the economic interpretation of each of the following null hypotheses:

(a. i) $\alpha_0 = \beta_0$.

(a. ii) $\alpha_1 = \beta_1$.

(a. iii) $\alpha_2 = \beta_2$.

Which of these hypotheses can be tested by using dummy variables? Where this can be done, explain exactly how you would do this.

Solution

a)

a. i) Constant term is the same in both regions.

a. ii) The effect of income is the same in both regions.

a. iii) The effect of size is the same in both regions.

Testing for each of the hypotheses involves using the dummy variable *REG* and testing the parameter on the dummy variable for significance. Let for example $REG_i = 0$ for the household in the South, while $REG_i = 1$ for the hoouseholds located in the North. Then the model allowed to test these hypoteses can be like this

$$CLOTH_i = \alpha_0 + \alpha_1 INC_i + \alpha_2 SIZE_i + \alpha_3 REG_i + \alpha_4 REG_i \cdot INC_i + \alpha_5 REG_i \cdot SIZE_i + u_{1i} \quad (1)$$

where $REG \cdot INC_i$ and $REG \cdot SIZE_i$ defined as it is shown here are slope dummies.

All hypotheses listed above (a.i, a.ii. and a.iii) are testable and corresponding tests are t-tests for coefficients $\hat{\alpha}_3$, $\hat{\alpha}_4$, $\hat{\alpha}_5$. Each of these tests is based on the assumption that location does not influence the other vaiables. For example, to test $H_0 : \alpha_3 = 0$ separately we assume that $\alpha_4 = \alpha_5 = 0$.

Marking

[5 mark] for the correct and complete explanations

(b) Consider the null hypotheses (a.i), (a.ii) and (a.iii) in (a) simultaneously. How this hypothesis can be tested by using dummy variables? Explain exactly how you would do this.

Solution

To test hypotheses (a.i), (a.ii) and (a.iii) in (a) simultaneously one should use F-test for three restrictions $H_0 : \alpha_3 = \alpha_4 = \alpha_5 = 0$. To do this one should run regression (1) in a) using full sample

$$CLOTH_i = \alpha_0 + \alpha_1 INC_i + \alpha_2 SIZE_i + \alpha_3 REG_i + \alpha_4 REG_i \cdot INC_i + \alpha_5 REG_i \cdot SIZE_i + u_{1i} \quad (1)$$

and evaluate the value of the sum of squared residuals RSS_U for the unrestricted model; then run restricted regression

$$CLOTH_i = \alpha_0 + \alpha_1 INC_i + \alpha_2 SIZE_i + \alpha_3 REG_i + u'_{1i} \quad (2)$$

and evaluate the value of the sum of squared residuals RSS_R for the restricted model; and then compare these values using F-test:

$$F = \frac{(RSS_R - RSS_U)/3}{RSS_U/(400-6)}$$

and appropriate critical values $F_{crit}^{\alpha\%}(3,394)$.

Marking

[5 mark] for the correct and complete explanations

(c) Consider now the null hypothesis (a.i) and (a.ii) in (a) simultaneously. How this hypothesis can be tested by using dummy variables? Explain exactly how you would do this.

Solution

To test hypotheses (a.i), (a.ii) and (a.iii) in (a) simultaneously one should use F-test for two restrictions $H_0 : \alpha_3 = \alpha_4 = 0$. To do this one should regression not using variable $REG_i \cdot SIZE_i$ (that is supposing that there are no difference in marginal effect of size between South and North)

$$CLOTH_i = \alpha_0 + \alpha_1 INC_i + \alpha_2 SIZE_i + \alpha_3 REG_i + \alpha_4 REG_i \cdot INC_i + u_{1i} \quad (3)$$

and evaluate the value of the sum of squared residuals RSS_U for the unrestricted model; then run restricted regression (the same as in b)

$$CLOTH_i = \alpha_0 + \alpha_1 INC_i + \alpha_2 SIZE_i + \alpha_3 REG_i + u'_{1i} \quad (2)$$

and evaluate the value of the sum of squared residuals RSS_R for the restricted model; and then compare these values using F-test:

$$F = \frac{(RSS_R - RSS_U)/2}{RSS_U /(400-5)}$$

and appropriate critical values $F_{crit}^{\alpha\%}(2,395)$.

Marking

[5 mark] for the correct and complete explanations

(d) The researcher wants to carry out more detailed analysis dividing the sample into four regions: North, South, East, West (suppose that each region is represented by the same number of families 100). She wants also to take into account the place of residence (urban or rural – both 50 families in each region). Explain how she could do this by using dummy variables.

Solution

As we have now four categories of location we need three dummy variables $SOUTH_i = 1$ for the households in the South, $EAST_i = 1$ for the households in the East, $WEST_i = 1$ for the households in the West, while households in the North are described by the combination of $SOUTH_i = 0$, $EAST_i = 0$, $WEST_i = 0$ (referebce category). To take into account the place of residence we need in additional dummy variable $URBAN_i = 1$ for the household in urban place, and correspondingly $URBAN_i = 0$ for the hoseholds in the country. The model could be like this

$$CLOTH_i = \gamma_0 + \gamma_1 INC_i + \gamma_2 SIZE_i + \gamma_3 SOUTH_i + \gamma_4 EAST_i + \gamma_5 WEST_i + \gamma_6 URBAN_i + u_{1i} \quad (4)$$

Reference category here is household in rural place in the North.

If we want to take into account an interaction between location and place of recidence, interactive dummies are introduced: $URBAN_i \cdot SOUTH_i$, $URBAN_i \cdot EAST_i$ and $URBAN_i \cdot WEST_i$. Of course to take into account the differences in marginal effects between categories the slope dummies could be also added like $SOUTH_i \cdot INC_i$ and so on. Using F-test for the full set of dummies including all slope dummies allows to decide whether there are any differences between locations and places of residence in the demand function.

MARKING

[5 mark] for the correct and complete explanations

(e) Explain how to do analysis in d) using Chow test. Explain in details.

Solution

We have here 4 locations each divided into two parts be the place of recidence: 8 subsamples total 50 observations each. To decide whether there are any differences in demand functions between subsamples one should run the original regression using whole sample

$$CLOTH_i = \delta_0 + \delta_1 INC_i + \delta_2 SIZE_i + u_T \qquad \text{(Total)}$$

and evaluate the value of the sum of squared residuals RSS_T, and then do the same using 8 subsamples,

$$CLOTH_i = \delta_{0k} + \delta_{1k} INC_{ik} + \delta_{2k} SIZE_{ik} + u_k \qquad \text{(Samples k=1,2,…,8)}$$

evaluating in each case RSS_k, $k = 1, 2, ..., 8$. Chow test uses F-statistic of the form

$$F = \frac{(RSS_T - RSS_1 - RSS_2 - ... - RSS_8)/(3 \cdot 8 - 3)}{(RSS_1 + RSS_2 + ... + RSS_8)/(400 - 3 \cdot 8)}$$

That could be compared with $F_{crit}^{\alpha\%}(21, 376)$ to decide whether there are any differences in demand functions between subsamples. Test is equivalent to the F-test for significance ot the full set of dummies including all slope dummies.

MARKING

[5 mark] for the correct and complete explanations

The International College of Economics and Finance
Econometrics – 2012-2013.
Mid-term exam 2013. March 29.

SUGGESTED SOLUTIONS AND MARKING

General instructions. Answer all questions of the problem 1 and any three of the problems 2-6. The weight of the obligatory problem 1 is 25% of the exam; three other questions add 25% each. You are advised to divide your time accordingly. Structure your answers in accordance with the structure of the questions. When testing hypotheses always state clearly null and alternative hypotheses provide critical value used for test, mentioning degrees of freedom and the significance level chosen for the test.

Part 1. (45 minutes).

1. Answer ALL FIVE SECTIONS of the first (obligatory) problem.

(a) Consider two equations

$$y_t = \alpha x_t + u_t \qquad (1)$$
$$y_t = \alpha x_t + \beta z_t + \varepsilon_t \qquad (2)$$

(a1) Let (1) be a **false model** while (2) be a **true model**. A researcher, using ordinary least squares (*OLS*), estimates α from the **false model**. Examine the properties for this *OLS* estimate of α.

Solution

$$\hat{\alpha}_{OLS} = \frac{\sum x_t y_t}{\sum x_t^2} = \frac{\sum x_t(\alpha x_t + \beta z_t + e_t)}{\sum x_t^2} = \alpha + \frac{\beta \sum x_t z_t}{\sum x_t^2} + \frac{\sum x_t e_t}{\sum x_t^2}.$$

$$E[\hat{\alpha}_{OLS}] = \alpha + \beta E\left[\frac{\sum x_t z_t}{\sum x_t^2}\right].$$

So $\hat{\alpha}_{OLS}$ is a biased estimator of α except in the case when x_t and z_t are **orthogonal** to each other ($\sum x_t z_t = 0$). Bias is $\beta \frac{\sum x_t z_t}{\sum x_t^2}$. Direction of the bias will depend upon the signs of β and $\sum x_t z_t$. Bias will disappear if x_t and z_t are **orthogonal** to each other.

Marking
[3 marks] for the correct and complete answer.

(a2) Let now (2) be a **false model** while (1) be a **true model**. A researcher, using ordinary least squares (*OLS*), estimates α from the **false model**. Examine the properties for this *OLS* estimate of α.

Solution

There is no bias, but the estimator $\hat{\alpha}_{OLS}$ is not efficient now, as more parameters are estimated therefore, the number of degrees of freedom is reduced. If correlation between x_t and z_t is r then standard error of $\hat{\alpha}_{OLS}$ is multiplied by $\frac{1}{\sqrt{1-r^2}}$ herefore it is possible that the estimated coefficient will be insignificant, estimated value of the coefficient is unpredictable, and it can even change sign. In the limiting case of perfect correlation of x_t and z_t the estimation is impossible (perfect multicollinearity).

Marking
[2 marks] for the correct and complete answer.
Total 5 marks for a)]

(b) A continuous random variable X has uniform distribution on the interval $[\alpha;\ \beta]$ where α and β are unknown. It means that the probability that X could take the value belonging to any interval $\Delta X \subseteq [\alpha;\ \beta]$ is proportional to the lengts of ΔX but does not depend of the position of $\Delta X \subseteq [\alpha;\ \beta]$, while the probability of taking any value outside the interval $[\alpha;\ \beta]$ is zero.

Suppose that a sample of n observations $x_1, x_2, ...; x_n \in [\alpha;\ \beta]$ is obtained. Derive the maximum likelihood estimators $\hat{\alpha}_{MLE}$ and $\hat{\beta}_{MLE}$ of parameters α and β. Let a sample consists of 5 observations $3.3;\ 2.7;\ 5.1;\ 0.7;\ 4.4$. Using this sample evaluate $\hat{\alpha}_{MLE}$ and $\hat{\beta}_{MLE}$.

Solution

b) Note that if X has uniform distribution on the interval $[\alpha;\ \beta]$ then probability density function is defined as $f_X(x) = \begin{cases} C \text{ if } x \in [\alpha;\ \beta] \\ 0 \text{ if } x \notin [\alpha;\ \beta] \end{cases}$. The value of the constant could be found from the condition $\int_{-\infty}^{+\infty} f_X(x)dx = \int_{\alpha}^{\beta} Cdx = 1$, so $C = \dfrac{1}{\beta - \alpha}$. So likelihood function is

$L(\alpha, \beta \,|\, x_1, ..., x_n) = f(\alpha, \beta \,|\, x_1) \cdot f(\alpha, \beta \,|\, x_2) \cdot ... \cdot f(\alpha, \beta \,|\, x_n) = \dfrac{1}{\beta - \alpha} \cdot \dfrac{1}{\beta - \alpha} \cdot ... \cdot \dfrac{1}{\beta - \alpha} = \left(\dfrac{1}{\beta - \alpha}\right)^n$. To maximize this function we need to take the values of α and β as close as possible to each other. But $x_1, x_2, ...; x_n \in [\alpha;\ \beta]$, so $\alpha \le \min x_i$, $\beta \ge \max x_i$, therefore $\hat{\alpha}_{MLE} = \min x_i$, $\hat{\beta}_{MLE} = \max x_i$.

Using the sample $3.3;\ 2.7;\ 5.1;\ 0.7;\ 4.4$ we find $\hat{\alpha}_{MLE} = 0.7$, $\hat{\beta}_{MLE} = 5.1$.

Marking

[5 mark] for the correct description of the Likelihood Function and sufficient explanations

(c) What is meant by a stationary time series? Under what conditions is the series generated by $X_t = \theta X_{t-1} + u_t$ stationary? Explain your answer.

Solution

c) A time series is stationary (weakly) if the mean, variance and autocovariance of the series are independent of time. Autocovariance may depend on length of the lag.

In this example $X_t = \theta X_{t-1} + u_t$; $t = 1, 2, ..., T$, where $E(u_t) = 0$; $\text{Var}(u_t) = \sigma^2$ and $E(u_s u_t) = 0$ for all s and t, $s \neq t$. Assume X_0 is fixed. We can write:

$$t = 1:\ X_1 = \theta X_0 + u_1,$$

$$t = 2:\ X_2 = \theta X_1 + u_2 = \theta(\theta X_0 + u_1) + u_2 = \theta^2 X_0 + \theta u_1 + u_2,$$

$$t = 3:\ X_3 = \theta X_2 + u_3 = \theta^3 X_0 + \theta^2 u_1 + \theta u_2 + u_3,$$

$$\vdots$$

Doing these recursive substitutions, we can write:

$$X_t = \theta X_{t-1} + u_t = \theta^t X_0 + u_t + \theta u_{t-1} + ... + \theta^{t-1} u_1.$$

Therefore:

$$E(X_t) = E(\theta^t X_0 + u_t + \theta u_{t-1} + ... + \theta^{t-1} u_1) = \theta^t X_0,$$

$$\text{Var}(X_t) = \text{Var}(\theta^t X_0 + u_t + \theta u_{t-1} + ... + \theta^{t-1} u_1) =$$

$$= \sigma_\varepsilon^2 (1 + \theta^2 + \theta^4 + ... + \theta^{2(t-1)}) = \sigma^2 \sum_{s=0}^{t-1} \theta^{2s}.$$

If $\theta \ge 1$ then for large 't', it is easy to see that $E(X_t) \to \infty$ and $\text{Var}(X_t) \to \infty$. So the variable X_t is non-stationary and standard analysis test statistics are not valid as these processes assume stationarity and finite mean and variance for the random variable. If this is the case then the variable is growing at the exponential rate which is rare for economic variables.

If $|\theta|<1$, then for large '*t*':

$$E(X_t)=\theta^t X_0=0,$$

and

$$\mathrm{Var}(X_t)=\sigma^2\sum_{s=0}^{t-1}\theta^{2s}=\sigma^2(1+\theta^2+\theta^4+...)=\frac{\sigma^2}{1-\theta^2},$$

which is a constant.
For large *t*, covariance is given by:

$$\mathrm{Cov}(X_t,X_{t-s})=E[X_t-E(X_t)][X_{t-s}-E(X_{t-s})]=E[X_tX_{t-s}]=$$
$$=E[u_t+\theta u_{t-1}+\theta^2u_{t-2}+...+\theta^{t-1}u_1][u_{t-s}+\theta u_{t-s-1}+...+\theta^{t-1}u_{-s}]=$$
$$=\theta^s(1+\theta^2+\theta^4+...)\sigma^2=\frac{\theta^s\sigma^2}{1-\theta^2},$$

which depends only upon the value of 's'. Therefore if $-1<\theta<1$. Hence the variable X_t is stationary if $|\theta|<1$.

Marking
[5 mark] for the correct explanations

(d) Let the regression equation be

$$Y_t=\alpha+\beta X_t+u_t;\ t=1,2,...,T.$$

where $E(u_t)=0$, $E(u_su_t)=\sigma^2X_t^2$ if $s=t$ and $E(u_su_t)=0$ if $s\neq t$.
Explain how you would estimate this equation to obtain best linear unbiased estimates of α and β.

Solution
d) Disturbance term is heteroscedastic, hence *OLS* estimators will be inefficient. Use Weighted Least Squares (*WLS*). Transform the model by dividing by X_t.

$$\frac{Y_t}{X_t}=\frac{\alpha}{X_t}+\beta+\frac{u_t}{X_t}\quad(1),$$

$$E\left[\frac{u_t}{X_t}\right]=0,$$

$$V\left(\frac{u_t}{X_t}\right)=\frac{\sigma^2X_t^2}{X_t^2}=\sigma^2,$$

$$E\left[\frac{u_su_t}{X_sX_t}\right]=0.$$

The transformed disturbance term in (1) has all the assumptions required for *OLS* estimators to be BLUE. Hence applying *OLS* to (1), $\hat{\alpha}$ and $\hat{\beta}$ can be obtained which will be BLUE. These are *WLS* estimators.
Marking
[5 mark] for the correct and complete explanations

(e) Suppose that the time series (for 36 consecutive years) DOC (expenditures of US citizens on doctors), DENT (expenditures on dentists), OPHT (expenditures on ophthalmologists) and PHAR (expenditures on drug preparations) form a panel (as observations for particular units of some general type of good (MED – medical services). The researcher is interested in evaluation of income elasticity and relative price elasticity on the base of the model

$$LOG(MED_{it})=\beta_1+\beta_2\cdot\log(DPI_t)+\beta_3\cdot\log(PRMED)_{it}+\varepsilon_t$$

The researcher wants to choose the best model. In each case describe briefly the test used, state the null hypothesis, characterize test statistics and indicate degrees of freedom?

Describe how the researcher could choose between fixed effects models and random effects models. The results of the test are

Hausman Test			
Test Summary	Statistic	d.f.	Prob.
Cross-section random	9.124375	2	0.0104

What model would you choose from this test? What are benefits and risks of your choice?
How to choose between random effects model and pooled regression?

Solution

e) The main question in the panel data analysis is the problem of the origin of unobserved heterogeneity $\sum_{p=1}^{s} \gamma_p Z_{pi} = \alpha_i$ in the model of the type

$$Y_{it} = \beta_0 + \sum_{j=1}^{k} \beta_j X_{jit} + \sum_{p=1}^{s} \gamma_p Z_{pi} + \varepsilon_{it} \qquad (1).$$

Fixed effect approach assigns it to the fixed characteristics of individual elements, while random effect approach assigns it to the random factors. For application of the random effect approach one should be possible to treat each of the unobserved Z_p variables as being drawn randomly from the same distribution. Moreover it is supposed that the Z_p variables are distributed independently of all of the X_j variables.
To decide which approach is more appropriate one should use Durbin–Wu–Hausman (DWH) test, that uses a statistic based on chi-square distribution with the number of degrees of freedom equal to the number of explanatory variables X_j in the equation (1) (equal to 2 in our case: $\log(DPI_t)$ and $\log(PRMED)$). The DWH test determines whether the estimates of the coefficients, taken as a group, are significantly different in the two regressions. The null hypothesis: the α_i are distributed independently of the X_j. In our case the null hypothesis is rejected at 5% level (as p-value=0.0104 < 0.05).

The additional argument in favor of fixed effect model is that four branches of medicine represented in the sample hardly be considered as random representative of medicine in general.

If the null hypothesis is correct, both random effects and fixed effects are consistent, but fixed effects will be inefficient because, it involves estimating an unnecessary set of dummy variable coefficients.
If the null hypothesis is correct, random effects is more attractive because observed characteristics that remain constant for each individual are retained in the regression model. With random effects estimation we do not lose n degrees of freedom, as is the case with fixed effects.
If the null hypothesis is false, the random effects estimates will be subject to unobserved heterogeneity bias and will therefore differ systematically from the fixed effects estimates.

If fixed effect approach or random effect approach is chosen there is always a possibility that there is no unobserved heterogeneity at all and one can apply pooled regression approach. To discriminate these cases there are some tests for this purpose, for example Breush-Pagan test based on Lagrange Multiplier approach. It uses also some chi-square distribution with $df = 1$ under H_0 of the absence of random effects.

Marking

[5 mark] for the correct and complete explanations

Part 2. (2 hours 15 minutes). Answer any three of the five problems (2-6).

2. Data on household expenditure on clothing (*CLOTH*) in dollars is obtained from a survey of 400 households in a two week period. In addition data on household income (*INC*) in dollars, the location of households (REGION) and the size of the family (*SIZE*) in number of persons was collected. An econometrician wants to test the hypothesis that the effects of income and family size are different in the North of the country as compared to the South. From a variable which denotes the region of residence of the household the data is divided into two sets, one for the North (200 households) and one for the South (200 households). The equations to be estimated are:

	REGION
$CLOTH_i = \alpha_0 + \alpha_1 INC_i + \alpha_2 SIZE_i + u_{1i}$	(North)
$CLOTH_i = \beta_0 + \beta_1 INC_i + \beta_2 SIZE_i + u_{2i}$	(South)

Where u_{1i} and u_{2i} are unobserved disturbances.

(a) Describe the economic interpretation of each of the following null hypotheses:

(a. i) $\alpha_0 = \beta_0$.

(a. ii) $\alpha_1 = \beta_1$.

(a. iii) $\alpha_2 = \beta_2$.

Which of these hypotheses can be tested by using dummy variables? Where this can be done, explain exactly how you would do this.

Solution

a)

a. i) Constant term is the same in both regions.

a. ii) The effect of income is the same in both regions.

a. iii) The effect of size is the same in both regions.

Testing for each of the hypotheses involves using the dummy variable *REG* and testing the parameter on the dummy variable for significance. Let for example $REG_i = 0$ for the household in the South, while $REG_i = 1$ for the hoouseholds located in the North. Then the model allowed to test these hypoteses can be like this

$$CLOTH_i = \alpha_0 + \alpha_1 INC_i + \alpha_2 SIZE_i + \alpha_3 REG_i + \alpha_4 REG_i \cdot INC_i + \alpha_5 REG_i \cdot SIZE_i + u_{1i} \quad (1)$$

where $REG \cdot INC_i$ and $REG \cdot SIZE_i$ defined as it is shown here are slope dummies.

All hypotheses listed above (a.i, a.ii. and a.iii) are testable and corresponding tests are t-tests for coefficients $\hat{\alpha}_3$, $\hat{\alpha}_4$, $\hat{\alpha}_5$. Each of these tests is based on the assumption that location does not influence the other vaiables. For example, to test $H_0: \alpha_3 = 0$ separately we assume that $\alpha_4 = \alpha_5 = 0$.

Marking

[5 mark] for the correct and complete explanations

(b) Consider the null hypotheses (a.i), (a.ii) and (a.iii) in (a) simultaneously. How this hypothesis can be tested by using dummy variables? Explain exactly how you would do this.

Solution

To test hypotheses (a.i), (a.ii) and (a.iii) in (a) simultaneously one should use F-test for three restrictions $H_0: \alpha_3 = \alpha_4 = \alpha_5 = 0$. To do this one should run regression (1) in a) using full sample

$$CLOTH_i = \alpha_0 + \alpha_1 INC_i + \alpha_2 SIZE_i + \alpha_3 REG_i + \alpha_4 REG_i \cdot INC_i + \alpha_5 REG_i \cdot SIZE_i + u_{1i} \quad (1)$$

and evaluate the value of the sum of squared residuals RSS_U for the unrestricted model; then run restricted regression

$$CLOTH_i = \alpha_0 + \alpha_1 INC_i + \alpha_2 SIZE_i + \alpha_3 REG_i + u'_{1i} \quad (2)$$

and evaluate the value of the sum of squared residuals RSS_R for the restricted model; and then compare these values using F-test:

$$F = \frac{(RSS_R - RSS_U)/3}{RSS_U/(400-6)}$$

and appropriate critical values $F_{crit}^{\alpha\%}(3,394)$.

Marking
[5 mark] for the correct and complete explanations

(c) Consider now the null hypothesis (a.i) and (a.ii) in (a) simultaneously. How this hypothesis can be tested by using dummy variables? Explain exactly how you would do this.

Solution
To test hypotheses (a.i), (a.ii) and (a.iii) in (a) simultaneously one should use F-test for two restrictions $H_0 : \alpha_3 = \alpha_4 = 0$. To do this one should regression not using variable $REG_i \cdot SIZE_i$ (that is supposing that there are no difference in marginal effect of size between South and North)

$$CLOTH_i = \alpha_0 + \alpha_1 INC_i + \alpha_2 SIZE_i + \alpha_3 REG_i + \alpha_4 REG_i \cdot INC_i + u_{1i} \quad (3)$$

and evaluate the value of the sum of squared residuals RSS_U for the unrestricted model; then run restricted regression (the same as in b)

$$CLOTH_i = \alpha_0 + \alpha_1 INC_i + \alpha_2 SIZE_i + \alpha_3 REG_i + u'_{1i} \quad (2)$$

and evaluate the value of the sum of squared residuals RSS_R for the restricted model; and then compare these values using F-test:

$$F = \frac{(RSS_R - RSS_U)/2}{RSS_U/(400-5)}$$

and appropriate critical values $F_{crit}^{\alpha\%}(2,395)$.

Marking
[5 mark] for the correct and complete explanations

(d) The researcher wants to carry out more detailed analysis dividing the sample into four regions: North, South, East, West (suppose that each region is represented by the same number of families 100). She wants also to take into account the place of residence (urban or rural – both 50 families in each region). Explain how she could do this by using dummy variables.

Solution
As we have now four categories of location we need three dummy variables $SOUTH_i = 1$ for the households in the South, $EAST_i = 1$ for the households in the East, $WEST_i = 1$ for the households in the West, while households in the North are described by the combination of $SOUTH_i = 0$, $EAST_i = 0$, $WEST_i = 0$ (referebce category). To take into account the place of residence we need in additional dummy variable $URBAN_i = 1$ for the household in urban place, and correspondingly $URBAN_i = 0$ for the hoseholds in the country. The model could be like this

$$CLOTH_i = \gamma_0 + \gamma_1 INC_i + \gamma_2 SIZE_i + \gamma_3 SOUTH_i + \gamma_4 EAST_i + \gamma_5 WEST_i + \gamma_6 URBAN_i + u_{1i} \quad (4)$$

Reference category here is household in rural place in the North.
If we want to take into account an interaction between location and place of recidence, interactive dummies are introduced: $URBAN_i \cdot SOUTH_i$, $URBAN_i \cdot EAST_i$ and $URBAN_i \cdot WEST_i$. Of course to take into account the differences in marginal effects between categories the slope dummies could be also added like $SOUTH_i \cdot INC_i$ and so on. Using F-test for the full set of dummies including all slope dummies allows to decide whether there are any differences between locations and places of residence in the demand function.

MARKING
[5 mark] for the correct and complete explanations

(e) Explain how to do analysis in d) using Chow test. Explain in details.

Solution
We have here 4 locations each divided into two parts be the place of recidence: 8 subsamples total 50 observations each. To decide whether there are any differences in demand functions between subsamples one should run the original regression using whole sample

$$CLOTH_i = \delta_0 + \delta_1 INC_i + \delta_2 SIZE_i + u_T \qquad \text{(Total)}$$

and evaluate the value of the sum of squared residuals RSS_T, and then do the same using 8 subsamples,

$$CLOTH_i = \delta_{0k} + \delta_{1k} INC_{ik} + \delta_{2k} SIZE_{ik} + u_k \qquad \text{(Samples k=1,2,…,8)}$$

evaluating in each case RSS_k, $k = 1, 2, \ldots, 8$. Chow test uses F-statistic of the form

$$F = \frac{(RSS_T - RSS_1 - RSS_2 - \ldots - RSS_8)/(3\cdot 8 - 3)}{(RSS_1 + RSS_2 + \ldots + RSS_8)/(400 - 3\cdot 8)}$$

That could be compared with $F_{crit}^{\alpha\%}(21, 376)$ to decide whether there are any differences in demand functions between subsamples. Test is equivalent to the F-test for significance ot the full set of dummies including all slope dummies.
MARKING
[5 mark] for the correct and complete explanations

3. Consider a model

$$Q_t = \alpha + \beta P_t^* + \gamma Z_t + u_t\,;\; t = 1, 2, ..., T \qquad (1)$$

where Q_t is the supply of wheat from the farmers of a particular country, P_t is the price of wheat (£) and Z_t is a measure of rainfall in this country. These three variables are observed. Price expectations P_t^* are revised by $P_t^* = \lambda P_{t-1} + (1-\lambda)P_{t-1}^*$. u_t is a random error, such that $E(u_t) = 0$; $E(u_t^2) = \sigma^2$ and $E(u_s u_t) = 0$ if $s \neq t$ for all $s, t = 1, 2, ..., T$.

(a) Show that this model can be reduced to infinite distributed lags model of the type

$$Q_t = \alpha + \beta \sum_{j=0}^{\infty} \mu^j P_{t-j-1} + \gamma Z_t + \varepsilon_t \qquad (2)$$

Solution

a) Starting from $P_t^* = \lambda P_{t-1} + (1-\lambda)P_{t-1}^*$

one can substitute lagged expression $P_{t-1}^* = \lambda P_{t-2} + (1-\lambda)P_{t-2}^*$ into it and get

$P_t^* = \lambda P_{t-1} + (1-\lambda)\lambda P_{t-2} + (1-\lambda)^2 P_{t-2}^*$, then repeating this procedure

$P_t^* = \lambda P_{t-1} + (1-\lambda)\lambda P_{t-2} + (1-\lambda)^2 \lambda P_{t-3} + (1-\lambda)^3 P_{t-3}^*$ and so on until

$P_t^* = \lambda P_{t-1} + (1-\lambda)\lambda P_{t-2} + (1-\lambda)^2 \lambda P_{t-3} + ... + (1-\lambda)^{k-1}\lambda P_{t-k} + (1-\lambda)^k P_{t-k}^*$

When $k \to \infty$ supposing $(1-\lambda) < 1$ we get $(1-\lambda)^k P_{t-k}^* \to 0$ and

$Q_t = \alpha + \beta(\lambda P_{t-1} + (1-\lambda)\lambda P_{t-2} + (1-\lambda)^2 \lambda P_{t-3} + ... + (1-\lambda)^{k-1}\lambda P_{t-k} + ...) + \gamma Z_t + u_t$

Denoting $\mu^j = (1-\lambda)^j \lambda$ we get $Q_t = \alpha + \beta \sum_{j=0}^{\infty} \mu^j P_{t-j-1} + \gamma Z_t + u_t$

MARKING

[5 mark] for the correct and complete explanations

Comment: At the examination there were some minor misprints in the question that did not prevent the correct answer.

(b) How the models of the type (2) could be estimated?

Solution

b) Disturbance term of this model inherits all properties of the original model.

The problem of estimating is the presence of the infinite lags while any data is finite, so we should stop substitution at certain point and omitting unobservable term $(1-\lambda)^k P_{t-k}^*$ that is supposed small enough

$Q_t = \alpha + \beta(\lambda P_{t-1} + (1-\lambda)\lambda P_{t-2} + (1-\lambda)^2 \lambda P_{t-3} + ... + (1-\lambda)^{k-1}\lambda P_{t-k}) + \gamma Z_t + u_t$

The direct estimation of this model using OLS is impossible because of non-linearity in parameters and the conflicting estimates of coefficients, but it is possible using non-linear least squares method. As P_{t-m} for different m are correlated the main problem in this estimation is multicollinearity. Another problem is the loss of k degrees of freedom that makes estimators inefficient.

Alternative approach is grid search: guessing the value of λ from 0 to 1 one can evaluate the sums $A_t(\lambda) = \lambda P_{t-1} + (1-\lambda)\lambda P_{t-2} + (1-\lambda)^2 \lambda P_{t-3} + ... + (1-\lambda)^{k-1}\lambda P_{t-k}$ and then use obtained values to estimate equation $Q_t = \alpha + \beta A_t(\lambda) + \gamma Z_t + u_t$ and to find $RSS(\lambda)$ of the equation; then we choose equation with the minimum $RSS(\lambda)$. This eliminates the problem of multicollinearity but the problem of reduced degrees of freedom and so inefficiency of estimators still remains.

MARKING

[5 mark] for the correct and complete explanations

(c) Show that the infinite distributed lags model of the type (2) can be written in terms of a single lag P_{t-1}. What estimation problems may occur in this model (besides the problem of serial correlation which is discussed later in e)?

Solution

c)

Starting from $Q_t = \alpha + \beta\lambda P_{t-1} + \beta(1-\lambda)\lambda P_{t-2} + \beta(1-\lambda)^2\lambda P_{t-3} + ... + \beta(1-\lambda)^{k-1}\lambda P_{t-k} + ...) + \gamma Z_t + u_t$ one can lag it and then multiply by $(1-\lambda)$

$(1-\lambda)Q_{t-1} = \alpha(1-\lambda) + \beta(1-\lambda)\lambda P_{t-2} + \beta(1-\lambda)^2\lambda P_{t-3} + ... + \beta(1-\lambda)^{k-1}\lambda P_{t-k} + ...) + \gamma(1-\lambda)Z_{t-1} + (1-\lambda)u_{t-1}$

and subtract the second equation from the first one; we get

$Q_t - (1-\lambda)Q_{t-1} = \alpha - \alpha(1-\lambda) + \beta\lambda P_{t-1} + \gamma(\lambda Z_t - (1-\lambda)Z_{t-1}) + u_t - (1-\lambda)u_{t-1}$ or

$Q_t = \alpha\lambda + \beta\lambda P_{t-1} + \gamma(Z_t - (1-\lambda)Z_{t-1}) + (1-\lambda)Q_{t-1} + u_t - (1-\lambda)u_{t-1}$

This is ADL(1,1) type model and usually model of this type could be estimated using OLS. But there are still some conflicting coefficients (γZ_t, $\gamma(1-\lambda)Z_{t-1}$ and $(1-\lambda)Q_{t-1}$ terms) so NLS required.

The problem with this estimation is also the fact that *RHS* variable Q_{t-1} which is a function of u_{t-1} is correlated with the error term so *OLS* produces inconsistent parameter estimates.

MARKING

[5 mark] for the correct and complete explanations

(d) Derive the short run and long run effect of *P* on *Q*.

Solution

d) Strictly speaking there is no direct relationship between current wheat supply Q_t and current prices P_t so the short run effect is zero. If we consider the issue more generally, considering short-term effects of two adjacent time periods, according to

$Q_t = \alpha + \beta\lambda P_{t-1} + \beta(1-\lambda)\lambda P_{t-2} + \beta(1-\lambda)^2\lambda P_{t-3} + ... + \beta(1-\lambda)^{k-1}\lambda P_{t-k} + ...) + \gamma Z_t + u_t$

short run effect of prices is $\beta\lambda$. It is the product of β by the speed of adaptation λ in the adaptive expectations equation $P_t^* = \lambda P_{t-1} + (1-\lambda)P_{t-1}^*$.

To get long run effect one need to substitute all values of variables by their equilibrium values $Q_t = \overline{Q}$, $P_t = \overline{P}$, $Z_t = \overline{Z}$ for all t.

$\overline{Q} = \alpha + \beta\lambda\overline{P} + \beta(1-\lambda)\lambda\overline{P} + \beta(1-\lambda)^2\lambda\overline{P} + ... + \beta(1-\lambda)^{k-1}\lambda\overline{P} + ...) + \gamma\overline{Z}$ or

$$\overline{Q} = \alpha + \beta\lambda\overline{P}(1 + (1-\lambda) + (1-\lambda)^2 + ... + (1-\lambda)^{k-1} + ...) + \gamma\overline{Z} = \alpha + \beta\lambda\overline{P}\frac{1}{1-(1-\lambda)} + \gamma\overline{Z} = \alpha + \beta\overline{P} + \gamma\overline{Z}$$

So $\overline{Q} = \alpha + \beta\overline{P} + \gamma\overline{Z}$, and long run effect of prices is β which is greater than short run effect $\beta\lambda$ ($0 < \lambda < 1$)

Alternative approach to answer this question is to use the ADL(1,1) form of the same model

$Q_t = \alpha\lambda + \beta\lambda P_{t-1} + \gamma(Z_t - (1-\lambda)Z_{t-1}) + (1-\lambda)Q_{t-1} + u_t - (1-\lambda)u_{t-1}$

Short run effect here is the same $\beta\lambda$. For the long run effect doing the same substitution we obtain

$\overline{Q} = \alpha\lambda + \beta\lambda\overline{P} + \gamma(\overline{Z} - (1-\lambda)\overline{Z}) + (1-\lambda)\overline{Q} = \alpha\lambda + \beta\lambda\overline{P} + \gamma\lambda\overline{Z} + (1-\lambda)\overline{Q}$

Moving $(1-\lambda)\overline{Q}$ to the left side of the equation we get $\lambda\overline{Q} = \alpha\lambda + \beta\lambda\overline{P} + \gamma\lambda\overline{Z}$, and dividing by λ we get $\overline{Q} = \alpha + \beta\overline{P} + \gamma\overline{Z}$ with the same conclusion that the long run effect of prices is β.

MARKING

[5 mark] for the correct and complete explanations

(e) Derive an econometric model in observable quantities directly using Koyck transformation (without using infinite lags). Discuss problems connected with serial correlation in the obtained model.

Solution and marking

e)

From $P_t^* = \lambda P_{t-1} + (1-\lambda)P_{t-1}^*$

$P_t^* - (1-\lambda)P_{t-1}^* = \lambda P_{t-1}$, hence from

$Q_t = \alpha + \beta P_t^* + \gamma Z_t + u_t$ and lagged form

$Q_{t-1} = \alpha + \beta P_{t-1}^* + \gamma Z_{t-1} + u_{t-1}$ multipliyng the lagged form by $(1-\lambda)$ we get

$$Q_t - (1-\lambda)Q_{t-1} = [\alpha - (1-\lambda)\alpha] + \beta(P_t^* - (1-\lambda)P_{t-1}^*) + \gamma(Z_t - (1-\lambda)Z_{t-1}) + u_t - (1-\lambda)u_{t-1} =$$
$$= \alpha\lambda + \beta\lambda P_{t-1} + \gamma Z_t - (1-\lambda)Z_{t-1} + u_t - (1-\lambda)u_{t-1}.$$

$$Q_t = \alpha\lambda + (1-\lambda)Q_{t-1} + \beta\lambda P_{t-1} + \gamma Z_t - (1-\lambda)Z_{t-1} + u_t - (1-\lambda)u_{t-1}.$$

Note that the error term $\varepsilon_t = u_t - (1-\lambda)u_{t-1}$ is serially correlated (negative autocorrelation of the moving average type). Indeed

$\varepsilon_t = u_t - (1-\lambda)u_{t-1}$ while $\varepsilon_{t-q1} = u_{t-1} - (1-\lambda)u_{t-2}$, so under assumptions of the original model

$$\mathrm{cov}(\varepsilon_t;\varepsilon_{t-1}) = \mathrm{cov}(u_t - (1-\lambda)u_{t-1}; u_{t-1} - (1-\lambda)u_{t-2}) = \mathrm{cov}(-(1-\lambda)u_{t-1}; u_{t-1}) = -(1-\lambda)\,\mathrm{cov}(u_{t-1}; u_{t-1}) = -(1-\lambda)\sigma_u^2$$

Therefore OLS estimators of coefficients will be inconsistent.

MARKING

[5 mark] for the correct and complete explanations

4. Consider a model

$$Y_i = \beta_0 + \beta_1 X_i + u_i;\ i = 1, 2, ..., n.$$

Where $Y_i = 1$ if the event takes place, $Y_i = 0$ otherwise and $E(u_i) = 0$.

(a) Explain clearly why there is a problem of heteroscedasticity by evaluating the variance of the disturbance term and examining its properties. Why is it a problem?

Solution

a) Model is:

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1, 2, ..., n$$

$Y_i = 1$ if the event takes place

$Y_i = 0$ otherwise

We note that

$$E[Y_i \mid X_i] = \beta_0 + \beta_1 X_i$$

$$\text{Also } E[Y_i \mid X_i] = 1 \cdot P(Y_i = 1) + 0 \cdot P(Y_i = 0) = P(Y_i = 1) = P_i$$

From (.72) and (.73), $E[Y_i \mid X_i] = \beta_0 + \beta_1 X_i = P_i$, hence we can interpret $E[Y_i \mid X_i] = \beta_0 + \beta_1 X_i$ as the probability that the event will occur, given X_i. As Y_i takes only two values 1 or 0, therefore u_i can take only two values: $1 - \beta_0 - \beta_1 X_i$ when $Y_i = 1$ and $-\beta_0 - \beta_1 X_i$ when $Y_i = 0$. Based on this we can write the probability distribution of u_i as

Y_i	u_i	$f(u_i)$
1	$1 - \beta_0 - \beta_1 X_i$	$\beta_0 + \beta_1 X_i$
0	$-\beta_0 - \beta_1 X_i$	$1 - \beta_0 - \beta_1 X_i$

As $E(u_i) = 0$ we can write $V(u_i)$ as

$$V(u_i) = E[u_i^2] = (1 - \beta_0 - \beta_1 X_i)^2(\beta_0 + \beta_1 X_i) + (-\beta_0 - \beta_1 X_i)^2(1 - \beta_0 - \beta_1 X_i) =$$

$$= (1 - \beta_0 - \beta_1 X_i)(\beta_0 + \beta_1 X_i) \times [(1 - \beta_0 - \beta_1 X_i) + (\beta_0 + \beta_1 X_i)] =$$

$$= (\beta_0 + \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i) = E[Y_i](1 - E[Y_i]) = P_i(1 - P_i)$$

$$\forall i = 1, 2, ..., n$$

Hence the disturbance term is heteroscedastic (P_i is different for different points of the sample). This will make *OLS* estimators inefficient. Weighted least squares can be used to obtain efficient estimators of β_0 and β_1.

Additional problems

i. In many cases the estimated probability $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ will be negative or greater than 1.

ii. As the distribution of the disturbance term only takes two values, it is not continuous. This implies that usual test statistics are invalid.

MARKING

[13 marks] for the correct and complete explanations

A researcher wants to examine the determinants of household decisions to buy alcohol. For this purpose she defines

$$Y = 1 \text{ if the household purchased alcohol}$$
$$= 0 \text{ otherwise.}$$

A random sample of households is available and the following **logit** estimates of the coefficients of variables were obtained:

Dependent variable: *Y*

Variables	Estimated Coefficients	Estimated Asymptotic Standard Errors
Income of the household (in ‘000s pounds)	158.39	48.86
square of income	-76.00	25.14
number of adults in the household	1.16	0.23
number of children in the household	0.49	0.12
1 if no worker in the household, 0 otherwise	-0.11	0.28
1 if head of the household is male, 0 if female	0.93	0.26
Constant	-83.94	23.77
$\log L$= -321.25	$\log L_0$= -416.01	

$\log L$ and $\log L_0$ are the log of the likelihood from the unrestricted model and log of the likelihood of the model where all the slope coefficients are restricted to zero, respectively.

(b) The researcher wanted to examine the estimated probability of the purchase of alcohol of a household which contains 2 adults, no children, 1 worker and head of the household is male. Obtain the estimated probability of purchase of alcohol by the kind of household described above if income is 0.9.

Solution

b)

$$\hat{Y}_i = -83.94 + 158.39(0.9) - 76(0.9)^2 + 1.16(2) + 0.93(1) = 0.301.$$

$$\hat{P}_i = P(Y_i = 1) = \frac{\exp(\hat{Y}_i)}{1 + \exp(\hat{Y}_i)} = \frac{\exp(0.301)}{1 + \exp(0.301)} = 0.5746.$$

MARKING

[3 marks] for the correct and complete explanations

(c) What would be the difference in result in **(b)** if the head of the household with 1 worker is female? Compare this difference with the marginal effect of gender estimated using derivatives

Solution

$$\hat{Y}_i = -83.94 + 158.39(0.9) - 76(0.9)^2 + 1.16(2) + 0.93(0) = -0.629.$$

$$\hat{P}_i = P(Y_i = 1) = \frac{\exp(\hat{Y}_i)}{1 + \exp(\hat{Y}_i)} = \frac{\exp(-0.629)}{1 + \exp(-0.629)} = 0.3477.$$

The probability of purchase of alhogol for male is greater by 0.5746-0.3477=0.2269.

Marginal effect of gender can be evaluated using derivatives and applying chain rule

$\frac{\partial Y}{\partial (gender)} = \frac{dp}{dZ} \cdot \beta_{gender} = \beta_{gender} \frac{e^{-z}}{(1+e^{-z})^2} = 0.93 \cdot \frac{e^{0.629}}{(1+e^{0.629})^2} = 0.2109$ that approximately equal to the value of 0.2269 found by direct calculation.

(d) Compare marginal effect obtained in c) with the maximum possible effect of gender.

Solution

As the derivative of the logistic function $\frac{dp}{dZ} = \frac{e^{-z}}{(1+e^{-z})^2}$ is symmetric ($\frac{e^{-z}}{(1+e^{-z})^2} = \frac{\frac{1}{e^z}}{(1+\frac{1}{e^z})^2} = \frac{e^z}{(1+e^z)^2}$) it takes its maximum value at the point $z = 0$ and this maximum value is $\frac{e^0}{(1+e^0)^2} = \frac{1}{(1+1)^2} = 0.25$. So the maximum marginal effect is $\frac{\partial Y}{\partial(gender)} = 0.25 \cdot \beta_{gender} = 0.25 \cdot 0.93 = 0.2325$. So in the household under consideration the marginal effect of gender on the purchase of alcohol is very close to its maximum value.

MARKING

[3 mark] for the correct and complete explanations

Comment: At the examination there was a misprint in the question so the students were instructed to ignore it. The corresponding points were distributed among other questions.

(e) Test the null hypothesis that coefficients of variables are all jointly equal to zero.

e As the model has been estimated by maximum likelihood, the *F*-test cannot be used. In these situations the likelihood ratio test is used. Likelihood ratio statistic is

$$LR = 2(\ln L - \ln L_0) = 2(-321.25 - (-416.01)) = 189.52.$$

Critical value of χ_6^2 at 5% level of significance is 12.592. Critical value of χ_6^2 at 1% level of significance is 16.812. Therefore reject H_0.

MARKING

[3 mark] for the correct and complete explanations

5. A student decided to investigate the behaviour of the exchange rates of USD, EURO and GBP with respect to Russian ruble in the period 01.01.2009 – 03.03.2013. He had downloaded 1034 daily observations from the Central Bank website (correctly deciding to neglect some irregularities due to missing data for national holidays).
First, the student tested the series for stationarity, doing Augmented Dickey-Fuller tests:

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RUR_USD)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RUR_USD(-1)	-0.010747	0.004494	-2.391431	0.0170
C	0.330455	0.137876	2.396755	0.0167

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RUR_EUR)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RUR_EUR(-1)	-0.007817	0.003828	-2.042050	0.0414
D(RUR_EUR(-1))	0.143295	0.031022	4.619186	0.0000
D(RUR_EUR(-2))	-0.082567	0.031037	-2.660280	0.0079
C	0.321210	0.158174	2.030741	0.0425

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RUR_GBP)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RUR_GBP(-1)	-0.022084	0.005978	-3.694129	0.0002
C	1.068431	0.288440	3.704171	0.0002

(a) Please explain the specification differences in the equations estimated and formulate the H_0 and H_1 hypotheses.

Solution

a) The first and the third specifications correspond to usual autoregression process $Y_t = \beta_1 + \beta_2 Y_{t-1} + \varepsilon_t$ and uses the equation $Y_t - Y_{t-1} = (\beta_2 - 1)Y_{t-1} + \varepsilon_t$. Null hypothesis is $H_0 : \beta_2 = 1$ (random walk – non stationary time series) againts $H_0 : -1 < \beta_2 < 1$ (stationary autoregressive process). The second case corresponds to the more complex model

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 Y_{t-2} + \beta_4 Y_{t-3} + \varepsilon_t .$$

Subtract Y_{t-1} from the both sides, add and subtract $\beta_3 Y_{t-1}$ and $\beta_4 Y_{t-1}$ on the right side

$$Y_t - Y_{t-1} = \beta_1 + \beta_2 Y_{t-1} + \beta_3 Y_{t-1} + \beta_4 Y_{t-1} - Y_{t-1} + \beta_3 Y_{t-2} - \beta_3 Y_{t-1} - \beta_4 Y_{t-1} + \beta_4 Y_{t-3} + \varepsilon_t .$$

Finally add and subtract $\beta_4 Y_{t-2}$ on the right side

$$Y_t - Y_{t-1} = \beta_1 + (\beta_2 + \beta_3 + \beta_4 - 1)Y_{t-1} + \beta_3 Y_{t-2} + \beta_4 Y_{t-2} - \beta_3 Y_{t-1} - \beta_4 Y_{t-1} - \beta_4 Y_{t-2} + \beta_4 Y_{t-3} + \varepsilon_t .$$

and come to the reparametrization

$$Y_t - Y_{t-1} = \beta_1 + (\beta_2 + \beta_3 + \beta_4 - 1)Y_{t-1} - (\beta_3 + \beta_4)(Y_{t-1} - Y_{t-2}) - \beta_4 (Y_{t-2} - Y_{t-3}) + \varepsilon_t .$$

or

$$\Delta Y_t = \beta_1 + (\beta_2 + \beta_3 + \beta_4 - 1)Y_{t-1} - (\beta_3 + \beta_4)\Delta Y_{t-1} - \beta_4 \Delta Y_{t-2} + \varepsilon_t .$$

For this process the Augmented Dickey-Fuller test is applicable
Null hypothesis is $H_0 : \beta_2 + \beta_3 + \beta_4 = 1$ (random walk – non stationary time series) againts $H_0 : -1 < \beta_2 + \beta_3 + \beta_4 < 1$ (stationary generalized autoregressive process). The optimum lag length could be chosen on the base of some information criteria (Akaike or Schwarz).

Marking

[5 marks] for correct explanation.

(b) Comment on the stationarity of each of three series, knowing that the ADF test statistics critical values are -2.86 for 5% level and -3.44 for 1% level.

Solution

b) The first process is certainly non-stationary as null hypothesis of random walk is not rejected at 5% level (-2.391431 > -2.86). The second one is also non-stationary at 5% level since -2.04>-2.86, but perhaps we do not reject the null hypothesis of non-stationarity due to small power of the test.

The third time series is stationary since -3.69<-3.44, and the null hypothesis of non-stationarity is rejected at the 1% level.

Comment: one should use critical values and not rely to p-value in the table as distribution of test statistic here is different from conventional t-statistics.The df is different in the model 2, but since it is large, the critical values are the same with the rounding used.

Marking

[5 marks] for correct explanation.

(c) Then the student decided to test if the series RUR_USD and RUR_EUR are cointegrated, and discovered that they are not. Please describe the concept of cointegration, and the procedure he did.

Solution

c) Two time series are cointegrated if

1) they are of the same order of integration say I(1) (first differences of the time series are stationary);

2) there is a linear combination of them that is stationary.

The second condition is equivalent to the following: the residuals of the regression of one of the time series on the other are stationary. So in practice one should show that RUR_USD and RUR_EUR are of the same order of integration, for example the first differences of these time series are stationary (this could be easily done with Dickey-Fuller test applied to the first differences). The one should regress RUR_USD on RUR_EUR (or vice versa) and investigate whether the residuals of this regression are stationary (standard procedure to be discussed, with the indication that the special critical values are taken because we teas the residuals instead of the disturbance term).

Marking

[5 marks] for correct explanation.

(d) Then the student decided to estimate the AR(1) and ADL(1,1) regressions and do the Common Factor test. For these two regressions the Residual Sums of Squares are 44.6953 and 44.66259 respectively. Calculate the test statistic (hint: $\log(1+\alpha) \approx \alpha$ for small α) and do the test. Which of two specifications would you choose, and why?

Solution

d) Let X_t and Y_t be two time series under consideration. If we assume that $Y_t = \beta_1 + \beta_2 X_t + u_t$ is AR(1) regression, then $u_t = \rho u_{t-1} + \varepsilon_t$. Using autoregressive transformation the autocorrelation could be removed from data.

It is also possible to transform the AR(1) model into specification

$Y_t = \beta_1(1-\rho) + \beta_2 X_t - \beta_2 \rho X_{t-1} + \rho Y_{t-1} + \varepsilon_t$ (the transformation to be described).

This is a special form for the model ADL(1,1) $Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 Y_{t-1} + \varepsilon_t$ under nonlinear restriction $\beta_3 = -\beta_2 \beta_4$. Testing this restriction (so called Common Factor test) uses chi-square statistic

$n \ln \dfrac{RSS_1}{RSS_0} = (1034-1) \cdot \ln \dfrac{44.6953}{44.66259} = 1033 \cdot \ln(1+0.00073238) \approx 1033 \cdot 0.00073238 = 0.7573$ what is less than critical value of 3.8415 (for the degrees of freedom equal to the number of restrictions =1). So the restriction is not rejected. As this restriction could be derived from assumption of autocorrelated disturbance term $u_t = \rho u_{t-1} + \varepsilon_t$ it means that ρ (coefficient of y_{t-1} could be interpreted as autocorrelation coefficient. The AR(1) model would be preferred to the ADL(1,1) model since the estimates in the model with valid restriction are more efficient.

Marking
[5 marks] for correct explanation.

(e) After that, the student reminded that he forgot to test the residuals for autocorrelation, and decided to test the ADL(1,1) model first. Explain why and how he could apply the Breusch-Godfrey LM test, with lag=2, for it.

Solution
e) The Breusch–Godfrey test allows to detect the residuals autocorrelation of the order higher than 1. It is a large samples test. It is applicable if the lagged dependent variable is one of the regressors which is important to the model considered. The most common realization of Breusch–Godfrey test involves the computation of the Lagrange multiplier statistic nR^2 when the residuals regression is fitted, n being the actual number of observations in the regression.
Let regression equation be $Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 Y_{t-1} + \varepsilon_t$. For the case of lag=2, we run this regression we find residuals e_t, and using them run auxiliary regression

$$\hat{e}_t == \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 Y_{t-1} + \rho_1 e_{t-1} + \rho_2 e_{t-2}.$$

Test statistic: nR^2, distributed as $\chi^2(2)$ when testing for second order autocorrelation. Alternatively, simple F test on coefficients of e_{t-1} and e_{t-2} , again with asymptotic validity. Here df=1034-1-2=1031 since 1 observation is lost in the ADL(1,1) model, and 2 in the auxiliary regression with lag=2.

Marking
[5 marks] for correct explanations.

6. A student took the data for 26 OECD countries, average for 2 time periods, on E (average annual percentage rate of growth of employment for country i during time period t) and G (average annual percentage rate of growth of GDP for country i during time period t). Having the panel of 52 observations, he decided to run the Fixed Effect regressions: Within Groups and First Differences.

(a) explain briefly the Fixed Effect approach and the two methods indicated**.**

Solution
a) The initial model for the panel data under consideration is

$$E_{it} = \beta_1 + \beta_2 G_{it} + \alpha_i + u_{it} \quad i = 1, ..., 26; t = 1,2.$$

α_i here is unobserved heterogeneity term that in Fixed Effect regression supposed to be a combined result of influence of some non-random factors $\alpha_i = \sum \gamma_j Z_{ij}$. To get rid of it one can use First Differences (FD) approach

$$E_{it} - E_{it-1} = \beta_1 + \beta_2 (G_{it} - G_{it-1}) + u_{it} - u_{it-1} \quad i = 1, ..., 26; t = 1.$$

This approach allows to eliminate unobservable α_i, at the cost of losing 26 degrees of freedom.
Alternative Within Groups (WG) approach is based on the idea of substitution instead of observations their deviations from the group means:

$$E_{it} - \bar{E}_i = \beta_1 + \beta_2 \left(G_{it} - \bar{G}_i\right) + u_{it} - \bar{u}_i$$

Marking
[5 marks] for correct explanation.

The student did generate the variables DE and DG (the first differences of E and G respectively), EEA and GGA (the deviations from the country average for E and G respectively.

After performing the estimations, the student got the following results:

Within Groups:
Dependent Variable: EEA
Included observations: 52

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GGA	0.690080	0.085139	8.105351	0.0000
R-squared	0.562970	Mean dependent var		0.000000
Sum squared resid	8.20669	S.D. dependent var		0.606796

First Differences
Dependent Variable: DE
Included observations: 26 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DG	0.690080	0.121603	5.674881	0.0000
R-squared	0.562952	Mean dependent var		-0.007692
Sum squared resid	16.41338	S.D. dependent var		1.225644

(b) Explain formally why the coefficient estimates in the Within Groups and First Differences regressions are the same.

Solution

b) The original equation is

$$E_{it} = \beta_1 + \beta_2 G_{it} + u_{it}$$

For two different years it gives

$$\begin{cases} E_{i1} = \beta_1 + \beta_2 G_{i1} + u_{i1} \\ E_{i2} = \beta_1 + \beta_2 G_{i2} + u_{i2} \end{cases}$$

WG group equations are

$$\begin{cases} E_{i1} - \dfrac{E_{i1}+E_{i2}}{2} = \beta_2\left(G_{i1} - \dfrac{G_{i1}+G_{i2}}{2}\right) + u_{i1} - \dfrac{u_{i1}+u_{i2}}{2} \\ E_{i2} - \dfrac{E_{i1}+E_{i2}}{2} = \beta_2\left(G_{i2} - \dfrac{G_{i1}+G_{i2}}{2}\right) + u_{i2} - \dfrac{u_{i1}+u_{i2}}{2} \end{cases}$$

or

$$\begin{cases} \dfrac{E_{i1}-E_{i2}}{2} = \beta_2\left(\dfrac{G_{i1}-G_{i2}}{2}\right) + \dfrac{u_{i1}-u_{i2}}{2} \\ \dfrac{E_{i2}-E_{i1}}{2} = \beta_2\left(\dfrac{G_{i2}-G_{i1}}{2}\right) + \dfrac{u_{i2}-u_{i1}}{2} \end{cases}$$

But this is just the same as in FD method

$$E_{i1} - E_{i2} = \beta_2(G_{i1} - G_{i2}) + u_{i1} - u_{i2}$$

Both methods minimize the same sum of squares

WG: $$\sum\left(\frac{E_{i1}-E_{i2}}{2} - \hat{\beta}_2\left(\frac{G_{i1}-G_{i2}}{2}\right)\right)^2 \to \min$$

FD: $$\sum\left((E_{i1}-E_{i2}) - \hat{\beta}_2(G_{i1}-G_{i2})\right)^2 \to \min$$

So the solutions to these problems (estimates of coefficients) are exactly the same.

Marking

[5 marks] for correct explanation.

(c) Explain the ratio of the sums of squared residuals in two models ($RSS_{FD}=2*RSS_{WG}$).

Solution

c) As it has been shown in b) within group method uses double quantity of observations and each observation is divided by 2. First (double quantity of observations) makes RSS_{WG} twice as big as RSS_{FD}, but the second (each observation is divided by 2) makes RSS_{WG} four times less. The final score is $RSS_{FD}=2*RSS_{WG}$.

(d) Derive the formula for the population variance for the LR model without constant $\sigma_{b_2}^2 = \dfrac{\sigma_u^2}{\sum_{j=1}^n X_j^2}$.

Solution

d)

$$\sigma_{b_2}^2 = \operatorname{var}\left(\frac{\sum_{j=1}^n X_j Y_j}{\sum_{j=1}^n X_j^2}\right) = \frac{\operatorname{var}\sum_{j=1}^n X_j(\beta_2 X_j + u_j)}{\left(\sum_{j=1}^n X_j^2\right)^2} = \frac{\operatorname{var}\beta_2 \sum_{j=1}^n X_l^2 + \operatorname{var}\sum_{j=1}^n X_j u_j}{\left(\sum_{j=1}^n X_j^2\right)^2} = \frac{0 + \sum_{j=1}^n X_j^2 \operatorname{var}(u_j)}{\left(\sum_{j=1}^n X_j^2\right)^2} =$$

$$= \frac{\sigma_u^2 \sum_{j=1}^n X_j^2}{\left(\sum_{j=1}^n X_j^2\right)^2} = \frac{\sigma_u^2}{\sum_{j=1}^n X_j^2}$$

Marking

[5 marks] for correct explanation.

(e) Basing on previous results, explain the ratio of the standard errors of coefficient estimates in two regressions.

Solution

e)

For the regression without intercept $\hat{\sigma}_u^2 = \dfrac{RSS}{n-1}$ so $s.e.(b_2) = \sqrt{\dfrac{\frac{RSS}{n-1}}{\sum_{j=1}^n X_j^2}}$.

Note that in $\sum_{j=1}^n X_j^2$ for WG method all observations are duplicated and twice as small as in FD method (see b), , so $\sum_{j=1}^n X_j^2$ for WG method is exactly a half of the corresponding sum for FD method. But the same relation is for RSS (see c): ($RSS_{FD}=2*RSS_{WG}$). It means that ratio $\dfrac{RSS}{\sum_{j=1}^n X_j^2}$ is the same for both methods (in FD there are 26 members of each sum, while in WG there are 52 members of each sum). So the ratio of standard errors is proportional to the square root of the inverse ratio for the numbers of observations minus one: $\dfrac{s.e.(b_2)_{FD}}{s.e.(b_2)_{WG}} = \sqrt{\dfrac{52-1}{26-1}} = \sqrt{\dfrac{51}{25}} = 1.428285686$. In fact if one multiply standard error of WG method by this factor we get exactly the standard error of FD method 0.085139*1.428285686=0.121603.

Marking

[5 marks] for correct explanation.

The International College of Economics and Finance
Econometrics – 2013-2014.
Midterm exam 2014. March 25.

General instructions. Candidates should answer EIGHT of the following TEN questions: ALL of the questions in Section A (8 marks each) and THREE questions from Section B (20 marks each). The weight of the Section A is 40% of the exam; three other questions add 20% each. You are advised to divide your time accordingly. Structure your answers in accordance with the structure of the questions. When testing hypotheses always state clearly null and alternative hypotheses provide critical value used for test, mentioning degrees of freedom and the significance level chosen for the test.

Section A (1 hour 15 minutes+10 minutes reading time)
Answer ALL FIVE problems of the first (obligatory) part (8 marks each).

1. Consider a regression model

$$y_i = \alpha x_i + u_i;\ i,\ j = 1, 2, \ldots, n$$

where $E(u_i) = 0$, $E(u_i^2) = \sigma^2 x_i^2$ and $E(u_i u_j) = 0$ if $i \neq j$ for all $i,\ j = 1, 2, \ldots, n$. x's are fixed.

(a) Derive the weighted least squares (WLS) estimator $\hat{\alpha}$, of α and also derive the variance of $\hat{\alpha}$.

(b) Is WLS estimator of α unbiased and consistent? Explain.

Solution: (a) WLS is applying OLS to the weighted equation $\frac{y_i}{x_i} = \alpha + \frac{u_i}{x_i}$.

Define $y_i^* = \frac{y_i}{x_i}$; $x_i^* = 1$; $u_i^* = \frac{u_i}{x_i}$. So WLS equation is $y_i^* = \alpha x_i^* + u_i^*$. The regression without constant can be estimated by $\hat{\alpha}^{OLS} = \frac{\sum y_i^* x_i^*}{\sum (x_i^*)^2} = \frac{\sum \frac{y_i}{x_i}\cdot 1}{\sum 1^2} = \frac{\sum \frac{y_i}{x_i}}{n}$

(b) As $\hat{\alpha}^{OLS} = \frac{1}{n}\sum \frac{y_i}{x_i} = \frac{1}{n}\sum \frac{\alpha x_i + u_i}{x_i} = \alpha + \frac{1}{n}\sum \frac{u_i}{x_i}$, it is not difficult to prove that the $\hat{\alpha}^{OLS}$ is unbiased:

$$E(\hat{\alpha}^{OLS}) = E\left(\alpha + \frac{1}{n}\sum \frac{u_i}{x_i}\right) = \alpha + \frac{1}{n}\sum \frac{Eu_i}{x_i} = \alpha$$

Now evaluate the variance of $\hat{\alpha}$:

$$E(\hat{\alpha}^{OLS} - \alpha)^2 = E\left(\frac{1}{n}\sum \frac{u_i}{x_i}\right)^2 = \frac{1}{n}E\left(\sum\left(\frac{u_i}{x_i}\right)^2\right) + \frac{1}{n^2}E\left(\sum_{j\neq i}^{n}\frac{u_j u_i}{x_j x_i}\right) = \frac{1}{n^2}\sum\left(\frac{\sigma^2 x_i^2}{x_i^2}\right) = \frac{\sigma^2}{n} \xrightarrow[n\to\infty]{} 0$$ so the sufficient condition for the consistency is satisfied.

2. Consider a regression model

$$Y_t = \beta_1 + \beta_2 X_t + u_t;\ t = 1, 2, ..., T$$

where $u_t = \varepsilon_t + \mu\varepsilon_{t-1}$ for all t and ε_t satisfies all reasonable assumtions $E(\varepsilon_t) = 0,$, $E(\varepsilon_t^2) = \sigma_\varepsilon^2$, $E(\varepsilon_s \varepsilon_t) = 0$ if $s \neq t$. The value of μ is supposed to be $0 < \mu < 1$ and small enough to neglect the values of degrees four (μ^4) and above.

(a) Show that this regression suffers from the autocorrelation of the moving average type.

(b) Show that using some iterative process similar iterative autoregressive transformation it is possible to get rid of autocorrelation, or at least to mitigate its consequences.

HINT: Substitute $u_t = \varepsilon_t + \mu\varepsilon_{t-1}$ into regression, then subtract from original regression lagged regression multiplied by μ. Repeat this transformation (with necessary modifications) at least twice and use assumptions on the value of μ.

Solution: (a) $u_t = \varepsilon_t + \mu\varepsilon_{t-1}$, so the same is true for $t-1$: $u_{t-1} = \varepsilon_{t-1} + \mu\varepsilon_{t-2}$. So

$$\operatorname{cov}(u_t, u_{t-1}) = \operatorname{cov}(\varepsilon_t + \mu\varepsilon_{t-1}, \varepsilon_{t-1} + \mu\varepsilon_{t-2}) = \operatorname{cov}(\varepsilon_t, \varepsilon_{t-1}) + \mu \operatorname{cov}(\varepsilon_t, \varepsilon_{t-2}) + \mu \operatorname{cov}(\varepsilon_{t-1}, \varepsilon_{t-1}) + \mu^2 \operatorname{cov}(\varepsilon_{t-1}, \varepsilon_{t-2})$$

$\operatorname{cov}(u_t, u_{t-1}) = 0 + 0 + \mu \operatorname{var}(\varepsilon_{t-1}) + 0 = \mu\sigma_\varepsilon^2$, as $E(\varepsilon_s \varepsilon_t) = 0$ if $s \neq t$.

(b) Substituting $u_t = \varepsilon_t + \mu\varepsilon_{t-1}$ into $Y_t = \beta_1 + \beta_2 X_t + u_t$ we get

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t + \mu\varepsilon_{t-1} \qquad (1)$$

The presence of ε_t is no problem, the problem is the presence of ε_t and ε_{t-1} simultaneously. Let's try to get rid of the latter. Lag the equation (1):

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + \varepsilon_{t-1} + \mu\varepsilon_{t-2} \qquad \text{(1: lag=1)}$$

and subtract from (1) obtained equation, multiplying both sides of it by μ

$$Y_t - \mu Y_{t-1} = \beta_1(1-\mu) + \beta_2(X_t - \mu X_{t-1}) + \varepsilon_t - \mu^2\varepsilon_{t-2} \qquad (2)$$

or, denoting $Y_t^{(1)} = Y_t - \mu Y_{t-1}$, $X_t^{(1)} = (X_t - \mu X_{t-1})$, and $\beta_1^{(1)} = \beta_1(1-\mu)$

$$Y_t^{(1)} = \beta_1^{(1)} + \beta_2 X_t^{(1)} + \varepsilon_t - \mu^2\varepsilon_{t-2} \qquad \text{(2a)}$$

Rewrite (2a) using lag equal to 2 periods

$$Y_{t-2}^{(1)} = \beta_1^{(1)} + \beta_2 X_{t-2}^{(1)} + \varepsilon_{t-2} - \mu^2\varepsilon_{t-4} \qquad \text{(2: lag=2)}$$

then multiply it by μ^2 and add obtained equation to (2):

$$Y_t^{(1)} + \mu^2 Y_{t-2}^{(1)} = \beta_1^{(1)}(1+\mu^2) + \beta_2(X_t^{(1)} + \mu^2 X_{t-2}^{(1)}) + \varepsilon_t - \mu^4\varepsilon_{t-4} \text{ (3),}$$

or, denoting $Y_t^{(2)} = Y_t^{(1)} + \mu^2 Y_{t-2}^{(1)}$, $X_t^{(2)} = X_t^{(1)} + \mu^2 X_{t-2}^{(1)}$, and $\beta_1^{(2)} = \beta_1^{(1)}(1+\mu^2)$,

$$Y_t^{(2)} = \beta_1^{(2)} + \beta_2 X_t^{(2)} + \varepsilon_t - \mu^4\varepsilon_{t-4} \qquad \text{(3a)}$$

As we can see from here the 'additional' disturbance term on the k^{th} stage of this repeating transformation is $-\mu^{2k}\varepsilon_{t-2k}$ and thus comes to zero. As before, $u_t = \varepsilon_t - \mu^{2k}\varepsilon_{t-2k}$ and $u_{t-2k} = \varepsilon_{t-2k} - \mu^{4k}\varepsilon_{t-4k}$ are correlated (having common term ε_{t-2k}): but this correlation

$$\operatorname{cov}(u_t, u_{t-2k}) = \operatorname{cov}(\varepsilon_t - \mu^{2k}\varepsilon_{t-2k},\ \ \varepsilon_{t-2k} - \mu^{4k}\varepsilon_{t-4k}) =$$

$$= \operatorname{cov}(\varepsilon_t, \varepsilon_{t-2k}) - \mu^{4k} \operatorname{cov}(\varepsilon_t, \varepsilon_{t-4k}) - \mu^{2k} \operatorname{cov}(\varepsilon_{t-2k}, \varepsilon_{t-2k}) + \mu^{6k} \operatorname{cov}(\varepsilon_{t-2k}, \varepsilon_{t-4k}) =$$

$$= 0 - 0 - \mu^{2k} \operatorname{var}(\varepsilon_{t-2k}) + 0 = -\mu^{2k}\sigma_\varepsilon^2$$

is weaker than initial correlation $\mu\sigma_\varepsilon^2$ (see **(a)**): ($0 < \mu < 1 \Rightarrow \mu^{2k}\sigma_\varepsilon^2 < \mu\sigma_\varepsilon^2$), and $\mu^{2k}\sigma_\varepsilon^2 \to 0$ with $k \to +\infty$
.

Note that the process of transformation can be organized in some different way, this influences only the speed of convergence in $\mu^m\sigma_\varepsilon^2 \to 0$ but does not change the conclusion.

Note also that this tranformation reduces the degrees of freedom in the finite samples.

3. Consider two linear regression models

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t \qquad (1)$$

$$y_t = \beta_1 + \beta_2 x_{2t} + u_t \qquad (2)$$

where the variables x_{2t}, x_{3t} are supposed to be non-stochastic, $u_t = \rho u_{t-1} + \varepsilon_t$, where $|\rho| < 1$ and $E(\varepsilon_t) = 0$; $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ and $E(\varepsilon_s \varepsilon_t) = 0$ for all *s* and *t.*

(a) Examine the properties of *OLS* estimate of β_2 using model (2), when the model (1) is **true**, while model (2) is **false**;

(b) Examine the properties of *OLS* estimate of β_2 using model (1), when the model (2) is **true**, while model (1) is **false**;

Solution: (a) Let model (1) is true. We show that in this case OLS estimator of β_2 in (2) will be biased.

$$\hat{\beta}_2 = \frac{\mathrm{Cov}(y_t, x_{2t})}{\mathrm{Var}(x_{2t})} = \frac{\mathrm{Cov}(\beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t, x_{2t})}{\mathrm{Var}(x_{2t})} = \frac{\mathrm{Cov}(\beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \rho u_{t-1} + \varepsilon_t, x_{2t})}{\mathrm{Var}(x_{2t})} =$$

$$= \beta_2 \frac{\mathrm{Cov}(x_{2t}, x_{2t})}{\mathrm{Var}(x_{2t})} + \beta_3 \frac{\mathrm{Cov}(x_{3t}, x_{2t})}{\mathrm{Var}(x_{2t})} + \rho \frac{\mathrm{Cov}(u_{t-1}, x_{2t})}{\mathrm{Var}(x_{2t})} + \frac{\mathrm{Cov}(\varepsilon_t, x_{2t})}{\mathrm{Var}(x_{2t})}.$$

Taking expectations we get $E\hat{\beta}_2 = \beta_2 + \beta_3 \frac{\mathrm{Cov}(x_{3t}, x_{2t})}{\mathrm{Var}(x_{2t})}$ as x_{2t}, x_{3t} are supposed to be non-stochastic, $\mathrm{Cov}(x_{2t}, x_{2t}) = \mathrm{Var}(x_{2t})$ and $E\,\mathrm{Cov}(u_{t-1}, x_{2t}) = \mathrm{cov}(u_{t-1}, x_{2t}) = 0$, $E\,\mathrm{Cov}(\varepsilon_t, x_{2t}) = \mathrm{cov}(\varepsilon_t, x_{2t}) = 0$.

The proof above is based on the following: from $E\varepsilon_t = 0$ follows $Eu_t = 0$: $Eu_t = \rho Eu_{t-1} + E\varepsilon_t = \rho Eu_t \Rightarrow Eu_t = \rho Eu_t$, as $\rho < 1 \Rightarrow Eu_t = 0$.

The sign and the magnitude of the bias depends on the values of β_3, $\mathrm{Cov}(x_{3t}, x_{2t})$ and $\mathrm{Var}(x_{2t}) > 0$.

The bias vanishes if x_{2t}, x_{3t} are orthogonal so $\mathrm{Cov}(x_{3t}, x_{2t}) = 0$.

(b) If now model (2) is true. There is no bias, but the estimators $\hat{\beta}_{2,OLS}$ is not efficient now, as more parameters are estimated therefore, the number of degrees of freedom is reduced. If correlation between x_{2t} and x_{3t} is r then standard error of $\hat{\beta}_{2,OLS}$ is multiplied by $\frac{1}{\sqrt{1-r^2}}$ herefore it is possible that the estimated coefficient will be insignificant, and even can change sign. In the limiting case of perfect correlation of x_t and z_t the estimation is impossible (perfect multicollinearity).

The presence of autocorrelated disturbance term does not change these conclusions. The estimates remain unbiased, and additional inefficiency emerges caused by autocorrelated disturbance term.

4. The researcher wants to test whether time series X_t is stationary using augmented Dickey-Fuller test including two additional lags: $X_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + u_t$ where u_t is distributed independently of X_t with zero mean and constant variance.

(i) without time trend;
(ii) including time trend;

(a) Derive Dickey-Fuller equation corresponding time series model under consideration in case **(i)**. What is the null hypothesis? What is the decision rule?

(b) What is Dickey-Fuller equation corresponding time series model in case **(ii)**. If the null hypothesis is rejected how to test the presence of time trend? If time trend is detected how to get rid of it?

Solution: (a) Starting from the equation above $X_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + u_t$ we first subtract the same term X_{t-1} from both sides:

$$X_t - X_{t-1} = \beta_0 + \beta_1 X_{t-1} - X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + u_t$$

Then add to and simultaneously subtract from the right side of equation the same terms $\beta_2 X_{t-1}$ and $\beta_3 X_{t-1}$:

$$X_t - X_{t-1} = \beta_0 + \beta_1 X_{t-1} - X_{t-1} + \beta_2 X_{t-1} - \beta_2 X_{t-1} + \beta_3 X_{t-1} - \beta_3 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + u_t$$

and rearrange terms taking out the common factors of the brackets

$$X_t - X_{t-1} = \beta_0 + (\beta_1 + \beta_2 + \beta_3 - 1) X_{t-1} - (\beta_2 + \beta_3) X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + u_t$$

Now we need to find a pair to the term $-(\beta_2 + \beta_3) X_{t-1}$ so add to and simultaneously subtract from the right side of equation the same terms $\beta_3 X_{t-2}$

$$X_t - X_{t-1} = \beta_0 + (\beta_1 + \beta_2 + \beta_3 - 1) X_{t-1} - (\beta_2 + \beta_3) X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-2} - \beta_3 X_{t-2} + \beta_3 X_{t-3} + u_t$$

And finally rearrange terms

$$X_t - X_{t-1} = \beta_0 + (\beta_1 + \beta_2 + \beta_3 - 1) X_{t-1} - (\beta_2 + \beta_3) X_{t-1} + (\beta_2 + \beta_3) X_{t-2} - \beta_3 X_{t-2} + \beta_3 X_{t-3} + u_t$$

$$X_t - X_{t-1} = \beta_0 + (\beta_1 + \beta_2 + \beta_3 - 1) X_{t-1} - (\beta_2 + \beta_3)(X_{t-1} - X_{t-2}) - \beta_3 (X_{t-2} - X_{t-3}) + u_t$$

or using differences

$$\Delta X_t = \beta_0 + (\beta_1 + \beta_2 + \beta_3 - 1) X_{t-1} - (\beta_2 + \beta_3) \Delta X_{t-1} - \beta_3 \Delta X_{t-2} + u_t$$

The same equation with the time trend is

$$\Delta X_t = \beta_0 + (\beta_1 + \beta_2 + \beta_3 - 1) X_{t-1} - (\beta_2 + \beta_3) \Delta X_{t-1} - \beta_3 \Delta X_{t-2} + \gamma t + u_t$$

In both cases (with and without time trend) the ADF test is based on the null hypothesis of no stationarity $H_0 : \beta_1 + \beta_2 + \beta_3 - 1 = 0$ against $H_a : \beta_1 + \beta_2 + \beta_3 - 1 < 0$ and uses special ADF statistics, with critical values calculated by computer. If the value of ADF less than critical null hypothesis is rejected.

(b) If the null hypothesis in (a) is rejected there is still the chance that the time series is time non-stationary, when a definite time trend violates the condition of constant expectation (one of conditions of so called weak stationarity). To test this it is sufficient to run conventional t-test for coefficient of time: $H_0 : \gamma = 0$ against $H_0 : \gamma \neq 0$.

If time trend is detected the simple procedure allows to get rid of it:

1) estimate equation $\hat{X}_t = \hat{\gamma}_0 + \hat{\gamma}_2 t$ and memorize its residuals $e_t = X_t - \hat{X}_t$;

2) use e_t instead of X_t in the model $X_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + u_t$ or another models under consideration.

Equivalent procedure is to include time variable explicitly into original equation

$$X_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \gamma t + u_t$$

5. Suppose that the time series $HOUS_t$ (Housing), WAT_t (Water), GAS_t (Gas), $FUEL_t$ (Fuel oil), KIT_t (Kitchen appliances), TAB_t (Tableware) form the panel (as observations for particular units of some general type of good ($GOOD_{it}$) related to maintaining a home). DPI_t is the disposable personal income and $PRGOOD_{it}$ is the relative price index for corresponding $GOOD_{it}$.
Let the model under investigation be

$$\log(GOOD_{it}) = \beta_1 + \beta_2 \log(DPI_t) + \beta_3 \log(PRGOOD_{it}) + u_{it}\,;\; i = 1, 2, ..., 50$$

The student runs three alternative approaches to the evaluation of this model: 1) pooled OLS regression, 2) fixed effects panel regression model, and 3) random effects panel regression model.

(a) Help the student to choose between different approaches: for each of three pairs of alternatives indicate corresponding test, null hypothesis, type of distribution of the test statistic, the number of degrees of freedom and the decision rule (which alternative is chosen if null hypothesis is rejected and which if it is not). What are advantages and risks of each choice?

(b) The model above is based on the assumption that the elasticities of different goods are the same. Suggest how to test this assumption. Give some details.

Solution: (a) Of course, only inexperienced student could consider three different tests together. In practice there is always a certain logic in the sequence of tests applied.
First the student is recommended to use Darbin-Wu-Hausman (DWH) test to choose between fixed and random effects. It is standard for majority of econometric computer programs and is based on using chi-square statistics with degrees of freedom equal to the number of variables in the equation under consideration (2 in our case) (as it compares estimates of coefficients obtained by two alternative models).
Under H_0 that there is no difference between coefficients obtained by two alternative models – fixed and random panel models (which means that unobserved heterogeneity α_i as a part of disturbance term, is not correlated with DPI_t and $PRGOOD_{it}$) both fixed effect and random effect models provide us with consistent estimates. We choose in this case random effect models as it retains in disturbance term all unobserved heterogeneity, there is no reduction of degrees of freedom typical for fixed effects models.
If H_0 is rejected, so there are essential differences between coefficients obtained using fized and random effects models, we choose fixed effects model, because rejecting of H_0 means that main assumption of independence of the disturbance term from regressors is violated so using random effects model we are under risk of getting inconsistent estimates of parameters.
So we have to suffice the fixed effects model that always gives consistent estimates.
Note that fixed effect model drops intercept (all factors that are constant in time) and includes a set of additional dummy variables (LSDV model) so it involves estimating an additional set of coefficients, that leads to reduction of degrees of freedom and so sometimes we observe more insignificant coefficients (for example in first difference or in within group versions of fixed effect models).

Further choice of the pair of model under consideration depends on the result of this test.
If the random effect model is chosen on the base of DWH test, it is possible that ordinary least squares is even better in efficiency if there is in fact no unobserved heterogeneity and so there no random effects at all. There are some tests for this purpose, for example Breush-Pagan test based on Lagrange Multiplier approach. It also uses chi-square distribution with degrees of freedom equal to 1 under H_0 of the absence of random effects.
If fixed effect model is chosen then conventional Chow test could be performed on the base of comparison of pooled regression with 6 separate regressions for $HOUS_t$, WAT_t, GAS_t, $FUEL_t$, KIT_t, and TAB_t under H_0 that there is no unobserved heterogeneity. If H_0 is rejected we choose fixed effect model, and we choose pooled regression if not.

(b) To answer this question is simpler in case of LDDV fixed effects model of the type

$$\log(GOOD_{it}) = \beta_2 \log(DPI_t) + \beta_3 \log(PRGOOD_{it}) + \sum_{i=1}^{6} \gamma_i D_i + u_{it}\,. \qquad (1)$$

where D_i are dummies corresponding different goods under consideration. To take into account that the elasticities of DPI_t and $PRGOOD_{it}$ could be different, it possible to run instead 6 different regressions for 6 goods ($i = 1, 2,..., 6$),

$$\log(GOOD_{it}) = \beta_1 + \beta_2 \log(DPI_t) + \beta_3 \log(PRGOOD_{it}) + u_{it}$$

Then perform an F-test using the values of RSS' of evaluated models:

$$F = \frac{(RSS_{LSDV} - \sum_{i=1}^{6} RSS_i)/(6\cdot 3 - (2+6))}{\sum_{i=1}^{6} RSS_i /(50 - 6\cdot 3)} = \frac{(RSS_{LSDV} - \sum_{i=1}^{6} RSS_i)/10}{\sum_{i=1}^{6} RSS_i /32}\,.$$

Here RSS_{LSDV} - the value of RSS for LSDV fixed effect model, $\sum_{i=1}^{6} RSS_i$ - sum of RSS' for separate regressions for different goods, the rule for degrees of freedom follows general principles of F-test and is clear from the formula above. If H_0 is not rejected there is no significant differences between elasticities.

Section B (1 hour 45 minutes+15 minutes reading time). Answer any three of the five problems (6-10).

6. A student under the guidance of her professor is studying the factors affecting the results of December 2013 exam in Econometrics. She is especially interested in interaction between the results of December multiple choice test (DM_i in %) and December free response (DF_i in %). She also believes that some other factors could influence the dependent variables such as the results of the October Econometrics exam (free response OF_i and multiple choice OM_i both in %), number of attended seminars A_i and average grade of submitted home assignments H_i (in %, zero if none submitted).

First she estimates two simultaneous equations

$$DF_i = \alpha_1 + \alpha_2 OF_i + \alpha_3 H_i + \alpha_4 A_i + u_i \quad (1)$$

$$DM_i = \beta_1 + \beta_2 OM_i + v_i \quad (2)$$

(the results of estimation for all equations see below).

The professor advised her to take into account the interaction between multiple choice and free response results, so variables DF_i and DM_i should be considered endogenous and all other variables exogenous. Following his advice she included variable DF_i instead of OM_i into second equation:

$$DF_i = \alpha_1 + \alpha_2 OF_i + \alpha_3 H_i + \alpha_4 A_i + u_i \quad (3)$$

$$DM_i = \beta_1 + \beta_2 DF_i + v_i \quad (4)$$

(a) Is there any difference between properties of the OLS estimators of these two systems of simultaneous equations (1-2) and (3-4)? Is OLS appropriate here? If not what can you recommend.

Then she decided to concentrate on the interaction between variables DF_i and DM_i, and estimated equations

$$DF_i = \alpha_1 + \alpha_2 DM_i + \alpha_3 OF_i + \alpha_4 H_i + \alpha_5 A_i + u_i \quad (5)$$

$$DM_i = \beta_1 + \beta_2 DF_i + v_i \quad (6)$$

(b) How this transformation changes the properties of estimators of the coefficients of equations? Provide proof of your conclusion.

(c) Describe the method that you can recommend for the estimation of the equation (6). Suggest some further transformation of the equations (5)-(6) producing consistent estimates for both equations.

	Regression (1)(3) *DF*		**Regression (2) *DM***		**Regression (4),(6) *DM***		**Regression (5) *DF***	
Variable	Coeff.	st.error	Coeff.	st.error	Coeff.	st.error	Coeff.	st.error
Const	4.39	1.85	26.67	3.80	29.46	2.78	0.16	2.37
DF					0.70	0.09		
DM							0.15	0.05
OF	0.36	0.05					0.30	0.05
OM			0.42	0.65				
H	0.18	0.04					0.17	0.04
A	0.46	1.92					0.35	0.24
R-squared	0.67		0.25		0.35		0.70	

Solution: (a) The variable DF_i in equation (1) depends only on the exogenous variables OF_i and H_i, the same is true for equation (2) where endogenous variable DM_i depends only on the exogenous variable OM_i, so the OLS estimates of both equation are unbiased, efficient and consistent (if all variables are considered as stochastic).

The only difference of the equations (3)-(4) from (1)-(2) is that the variable DF_i is included into second equation instead of OM_i. The estimates of equation (3) remain of course unbiased. As there is no circular dependence in the system of equations (3)-(4) the estimates of (4) are unbiased and consistent.

(b) As the dependent variable DF_i of the equation (5) is included in (6) as an explanatory variable and vice versa the dependent variable DM_i of the equation (6) is included in (5) as an explanatory variable, the circularity emerges that leads to the fact that the disturbance term u_i correlates with the explanatory variable DF_i in equation (5) and the disturbance term v_i correlates with the explanatory variable DM_i in equation (2).

(Detailed explanation: let '~' means 'correlates with', then from equation (5) $DF_i \sim u_i$ then from equation (6) $DM_i \sim DF_i$ and so $DM_i \sim u_i$, that means that Gauss-Markov conditions are violated for equation (5). The same could be shown for equation (2).)

It can be shown that they are inconsistent and so biased for the large samples but the proof of this requires more advanced technique.

The system of equations

$$DF_i = \alpha_1 + \alpha_2 DM_i + \alpha_3 OF_i + \alpha_4 H_i + \alpha_5 A_i + u_i \qquad (5)$$

$$DM_i = \beta_1 + \beta_2 DF_i + v_i \qquad (6)$$

is called to be in structural form.

Reduced form equation can be derived from here (we need only one of them).

$$DF_i = \alpha_1 + \alpha_2(\beta_1 + \beta_2 DF_i + v_i) + \alpha_3 OF_i + \alpha_4 H_i + \alpha_5 A_i + u_i$$

$$(1-\alpha_2\beta_2)DF_i = \alpha_1 + \alpha_2\beta_1 + \alpha_3 OF_i + \alpha_4 H_i + \alpha_5 A_i + u_i + \alpha_2 v_i$$

$$DF_i = \frac{1}{(1-\alpha_2\beta_2)}[\alpha_1 + \alpha_2\beta_1 + \alpha_3 OF_i + \alpha_4 H_i + \alpha_5 A_i + u_i + \alpha_2 v_i] \qquad (7)$$

Show now that OLS estimator of β_2 in (6) is inconsistent. The sketch of this proof (omitting unnecessary details, as it is important to establish only the fact of the bias for large samples) is like this

$$\hat{\beta}_2^{OLS} = \frac{\mathrm{Cov}(DM_i, DF_i)}{\mathrm{Var}(DF_i)} = \beta_2 + \frac{\mathrm{Cov}(v_i, DF_i)}{\mathrm{Var}(DF_i)}, \text{ so } \operatorname{plim} \hat{\beta}_2^{OLS} = \beta_2 + \frac{\operatorname{plim}\mathrm{Cov}(v_i, DF_i)}{\operatorname{plim}\mathrm{Var}(DF_i)} = \beta_2 + \frac{\mathrm{cov}(v_i, DF_i)}{\mathrm{var}(DF_i)}.$$

Variance $\mathrm{var}(DF_i)$ is in any case positive. From (7) DF_i is a linear combination of exogenous variables OF_i, H_i, A and disturbance terms u_i and v_i. As disturbance term v_i can be supposed to be uncorrelated with exogenous variables OF_i, H_i, A, then omitting all zero population covariances and constant common factor, and also using expression for the disturbance term from (7),

$\mathrm{cov}(v_i, DF_i) = \dfrac{\mathrm{cov}(v_i, u_i + \alpha_2 v_i)}{1-\alpha_2\beta_2} = \dfrac{\mathrm{cov}(v_i, u_i) + \alpha_2 \,\mathrm{var}(v_i)}{1-\alpha_2\beta_2}$. As disturbance terms can be supposed uncorrelated the fact of bias is proven, as $\mathrm{cov}(v_i, DF_i) = \dfrac{\alpha_2 \,\mathrm{var}(v_i)}{1-\alpha_2\beta_2} \neq 0$.

(c) Equation (6) is over-identified, as there are a lot of available instruments (OF_i, H_i, A) for endogenous variable DF_i in the right side of equation (6). In this case the most efficient way to get consistent estimate for β_2: TSLS - Two Stage Least Squares. At the first stage estimate equation (5)

$$DF_i = \alpha_1 + \alpha_2 DM_i + \alpha_3 OF_i + \alpha_4 H_i + \alpha_5 A_i + u_i \qquad (5)$$

and memorize the estimated values of $\hat{DF}_i$. At the second stage use these estimated values $\hat{DF}_i$ instead of DF_i to estimate parameters of the equation (6).

Another problem with the system (5)-(6) is that first equation is underidentified. So it is better to try redistribute explanatory variables among two equations. For example it is possible to move one explanatory variable, say A_i from equation (5) to equation (6) and get the system

$$DF_i = \alpha_1 + \alpha_2 DM_i + \alpha_3 OF_i + \alpha_4 H_i + u_i \qquad (5')$$

$$DM_i = \beta_1 + \beta_2 DF_i + \beta_3 A_i + v_i \qquad (6')$$

where equation (5') is now exactly identified while *6') is still over-identified.

Another suggestion is using some appropriate variables from outside the system, for example rating of students $R2_i$ made at the end of the second year of study:

$$DF_i = \alpha_1 + \alpha_2 DM_i + \alpha_3 OF_i + \alpha_4 H_i + \alpha_5 A_i + u_i \qquad (5')$$

$$DM_i = \beta_1 + \beta_2 DF_i + \beta_3 R2_i + v_i \qquad (6')$$

with the same results. There is a general rule how to choose between available instruments: they should correlate with the instrumented endogenous variable (in our case DM_i in the right side of the first equation) and be uncorrelated with disturbance term of both equations.

7. Recently, the Central Bank of Russian Federation revoked the licenses of several Russian banks, some other banks are considered as problematic and some of them are at risk of revocation of license in the near future. In this connection customers and entrepreneurs are interested in the estimating the state of the bank to select reliable banks. Since the complete information about the bank is not available, the problem arises of express analysis of financial and economic condition of the bank based on the values of a limited number of the most informative indicators of its performance. Preliminary expert analysis showed that such informative variables can be attributed as:

X_1 - percentage of the amount in foreign currency liabilities in the balance sheet of the bank;

X_2 - share of total loans in the amount of operating assets, and

X_3 - the ratio of the balance sheet profit to balance losses.

A survey was conducted of 500 of the largest (in terms of value of liabilities and capital), which resulted in the following data: ($X_{1i}, X_{2i}, X_{3i}, Y_i, i = 1,...,500$), where $Y_i = 1$, if the Central Bank attributes this bank to the category of problematic banks, and $Y_i = 0$ otherwise (this variable can be called a reliability index of the bank).

Suppose you are invited as a consultant to build the regression model Y_i as a function of X_{1i}, X_{2i}, X_{3i}.

(a) Explain why linear regression model $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ estimated using OLS is not appropriate in this case.

(b) Suggest some model and corresponding procedure that allow to estimate Y_i as a function of X_{1i}, X_{2i}, X_{3i}, providing the rationale of the proposed form of the function, and explaining why the obtained estimates have acceptable econometric properties. How on the basis of the proposed model to estimate that some factor, say, X_{3i}, is significant? On the basis of the same model how to estimate that a group of factors, for example, X_{1i}, X_{2i}, is significant

(c) Describe in detail the procedure for the econometric analysis of the proposed model on the basis of your original data. Suppose some characteristic, say X_{3i} - the ratio of the balance sheet profit to balance losses, of the bank under consideration i has changed while maintaining other explanatory variables unchanged. Is it possible to assess how it will affect the reliability index of the bank without a recalculation of the constructed function? Is it possible to evaluate maximum marginal effect of some factor, say X_{3i} (while keeping X_{1i}, X_{2i} fixed)?

Solution: (a) Y_i is a binary choice variable, that in any model of the type

$$Y_i = f(X_{1i}, X_{2i}, X_{3i}) + u_i$$

Y_i can be interpreted as probability that the bank will be considered as problematic.

The linear probability model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

Estimated by OLS, is not appropriate here as there are 4 main disadvantages of the linear probability model:

1) possible predicted values can be outside the range of probability [0; 1];

2) it assumes constant marginal effect of each factor;

3) heteroscedasticity is present caused by abnormal discontinuous disturbance term,

$$u_i = 1 - \beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \beta_3 X_{3i} \quad \text{if } Y_i = 1$$

$$u_i = -\beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i} - \beta_3 X_{3i} \quad \text{if } Y_i = 0$$

and so Gauss-Markov conditions are violated, so OLS produces inefficient estimates;

4) the distribution of the disturbance term is not normal, so the useful tests of coefficients are not valid.

(b) So the best solution is to construct the binary choice model

$$Y_i = F(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i)$$

where $F(z)=\dfrac{1}{1+e^{-z}}$ (logit model), or $F(z)=\dfrac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{z} e^{-\frac{x^2}{2}}dx$ (probit model).

These two model are quite similar in their properties, the most important of them are

1) predicted values $\hat{Y}_i$ belong to the range [0; 1];
2) the marginal effect is not constant, it is maximal for $z=0$ and is negligible for extreme values of z;
For estimation of the models the maximum likelihood estimation method is used that provides asymptotically efficient estimates.
The new system of tests based on the values of logarithmic likelihood function, is used instead of tests based on RSS in OLS estimation. To estimate that some factor, say, X_{3i}, is significant, it is sufficient to evaluate its asymptotic z-values (analogous conventional t-statistics) (provided by the software), and compare it with the critical values of standard normal distribution. All tests under consideration are asymptotic so the sample is supposed to be large enough (50 in our case).
To evaluate the significance of the group of two X_{1i}, X_{2i} (or more) variables we need some analog of F-test. It is so called LR-test based on the statistic $LR=2(\log L_U-\log L_R)$, where L_U is likelihood function of the unrestricted model, evaluated for the function including all variables

$$Y_i=F(\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\beta_3X_{3i}+u_i)$$

and L_R is the value of likelihood function evaluated for the restricted model ($H_0:\beta_1=\beta_2=0$):

$$Y_i=F(\beta_0+\beta_1X_{3i}+u_i).$$

LR-statistic has χ^2-distribution with the degrees of freedom equal to the number of restrictions (2 in our case).

(c) To evaluate the marginal effect of some factor say X_{3i} we should use the defferential of the function $Y_i=F(\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\beta_3X_{3i}+u_i)$. For logit model it is evaluated as $\dfrac{e^{-z}}{(1+e^{-z})^2}\cdot\beta_3 dX_{3i}$, where z is evaluated in the in the initial point (X_{1i},X_{2i},X_{3i}): $z=\hat{\beta}_0+\hat{\beta}_1X_{1i}+\hat{\beta}_2X_{2i}+\hat{\beta}_3X_{3i}$ and dX_{3i} is a small change of X_{3i}. For the probit model the marginal effect is evaluated as $\dfrac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\cdot\beta_1 dX_{3i}$.

As it is said above the maximum effect in both cases achieved for $z=0$, so to find the value of X_{3i} for which the maximum effect achieved under fixed X_{1i}, X_{2i}. To do this we have to solve the equation

$$\hat{\beta}_0+\hat{\beta}_1X_{1i}+\hat{\beta}_2X_{2i}+\hat{\beta}_3X_{3i}=0$$

for X_{3i}.

8. An econometrician having **quarterly data** for 12 years (plus current values - 49 observations total) believes that current total consumption expenditure C_t is dependent not only on current value of disposable personal income Y_t and current price index P_t, but also on the last **two** years values of disposable personal income Y_{t-k}. She estimates using OLS the equation:

$$\hat{C}_t = 99 + 0.9Y_t - 0.2Y_{t-1} - 0.4Y_{t-2} - 0.2Y_{t-3} - 0.1Y_{t-4} + 0.04Y_{t-5} + 0.1Y_{t-6} + 0.4Y_{t-7} - 0.02Y_{t-8} + 0.4P_t \quad R^2 = 0.99$$

(91) (0.32) (0.28) (0.29) (0.29) (0.28) (0.30) (0.33) (0.33) (0.30) (0.31)

(a) What econometric phenomena can be observed in the equation above? What econometric problems (if any) are likely connected with these phenomena?

Explain how you would test the hypothesis that consumption is dependent on disposable income for the last year only against the alternative hypothesis that consumption is dependent on the last two years. Give details of the information you need for this test.

(b) A colleague suggests that you should use an infinite lag model instead of the model above in **(a)**. How would you estimate this model

i) on the basis of Koyck distribution;

ii) on the basis of Koyck transformation?

What are benefits and disadvantages of these approaches (compare properties of OLS estimators for different models)?

(c) How the approach used in **(b)** is connected with adaptive expectation model?

Solution: (a) Assuming the data is quarterly then 2 years of data will give a model with 8 lags. The econometric problem is one of multicollinearity, lagged values are likely to be very similar to each other unless the variable is essentially random and this is unlikely to be true for disposable income which is going to be slowly changing. As the result we observe non-significant coefficients, many of them have the wrong sign. The equation is useless for analysis, but can be still well be used for the prediction (R-squared is very high). As lags cover only 2 years the omitted variable bias is also present.

The 2-year model will be

$$C_t = \beta_0 + \beta Y_t + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4} + \beta_5 Y_{t-5} + \beta_6 Y_{t-6} + \beta_7 Y_{t-7} + \beta_8 Y_{t-8} + \gamma P_t + u_t .$$

If the current value of consumption expenditure does not depend on the data the year before, the equation will be

$$C_t = \beta_0 + \beta Y_t + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4} + \gamma P_t + u_t .$$

with the restrictions $H_0 : \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$.

against the alternative that at least one of the β_s ($5 \le s \le 8$) are non-zero. Note that using lagged variables reduces the number of observation by 8.

This test can be accomplished by an *F*-test

$$F_{4,26} = \frac{(RSS_R - RSS_U)/4}{RSS_U/(45-8-11)},$$

where RSS_R is the restricted residual sum of squares under H_0 and RSS_U is the unrestricted residual sum of squares obtained by running the regression on all the lagged variables.

NOTE: The formula for the *F*-test of restrictions should be given. Degrees of freedom should be clearly specified.

(b) The infinite lagged variable model would be constructed with geometrically declining weights, for example

$$C_t = \beta_0 + \beta \sum_{j=0}^{\infty} \lambda^j Y_{t-j} + \gamma P_t + u_t$$

where $|\lambda|<1$. The advantage of this approach is that there are only three parameters to estimate but against this the weights of successive values of Y_t will be forced to decline geometrically which may not be appropriate.
There are several approaches to estimation of the parameters of this model. Nonlinear estimation implies that first we choose some value of λ and calculate $\sum_{j=0}^{\infty}\lambda^j Y_{t-j}$ (or its limited part on the base of available data). Then using OLS parameters β_0, β, γ are estimated. We should try many different variants of λ (grid search) to choose the best regression with least RSS (of standard error of regression).

Another more elegant approach is based of the transformation the above equation to include a lagged dependent variable. Note that if we lag the above equation

$$C_t = \beta_0 + \beta\sum_{j=0}^{\infty}\lambda^j Y_{t-j} + \gamma P_t + u_t$$

by one period

$$C_{t-1} = \beta_0 + \beta\sum_{j=0}^{\infty}\lambda^j Y_{t-j-1} + \gamma P_{t-1} + u_{t-1}$$

and multiply the resulting equation by λ we get

$$\lambda C_{t-1} = \lambda\beta_0 + \beta\sum_{j=0}^{\infty}\lambda^j Y_{t-j-1} + \gamma\lambda P_{t-1} + \lambda u_{t-1}$$

so by subtracting them

$$C_t - \lambda C_{t-1} = (1-\lambda)\beta_0 + \beta Y_t + \gamma P_t - \lambda\gamma P_{t-1} + u_t - \lambda u_{t-1}$$

So we come to the equation

$$C_t = (1-\lambda)\beta_0 + \beta Y_t + \gamma P_t - \lambda\gamma P_{t-1} + \lambda C_{t-1} + u_t - \lambda u_{t-1}$$

This is a restricted version of ADL(1, 1) model (the coefficient by P_{t-1} is minus product of coefficient γ of current prices P_t and the coefficient λ of the lagged dependent variable C_{t-1}). It cannot be estimated using OLS and require some nonlinear technique. The restriction should be tested using common factor test and it the restriction is not valid we should use more general ADL(1,1) model

$$C_t = (1-\lambda)\beta_0 + \beta Y_t + \gamma P_t + \delta P_{t-1} + \lambda C_{t-1} + u_t - \lambda u_{t-1}$$

that can be estimated using OLS.
Another problem with this model is that the error term $\varepsilon_t = u_t - (1-\lambda)u_{t-1}$ is serially correlated (negative autocorrelation of the moving average type). Indeed
$\varepsilon_t = u_t - (1-\lambda)u_{t-1}$ while $\varepsilon_{t-1} = u_{t-1} - (1-\lambda)u_{t-2}$, so under assumptions of the original model

$$\operatorname{cov}(\varepsilon_t;\varepsilon_{t-1}) = \operatorname{cov}(u_t - (1-\lambda)u_{t-1}; u_{t-1} - (1-\lambda)u_{t-2}) = \operatorname{cov}(-(1-\lambda)u_{t-1}; u_{t-1}) =$$
$$= -(1-\lambda)\operatorname{cov}(u_{t-1}; u_{t-1}) = -(1-\lambda)\sigma_u^2$$

The models above include lagged dependent variable therefore OLS estimators of coefficients will be inconsistent.

(c) If the level of current consumption is determined not by current value of disposable personal income but by the expectation of is its future value (and of current prices)

$$C_t = \alpha + \beta Y_{t+1}^* + \gamma P_t + u_t$$

we can additionally assume that

$$Y_{t+1}^* = \lambda Y_t + (1-\lambda)Y_t^*$$

or

$$Y_t^* - (1-\lambda)Y_{t-1}^* = \lambda Y_t,$$

hence from

$$C_t = \alpha + \beta Y_{t+1}^* + \gamma P_t + u_t$$

and its lagged form

$$C_{t-1} = \alpha + \beta Y_t^* + \gamma P_{t-1} + u_{t-1}$$

multipliyng the lagged form by $(1-\lambda)$ we get

$$C_t-(1-\lambda)C_{t-1}=[\alpha-(1-\lambda)\alpha]+\beta(Y_{t+1}^{*}-(1-\lambda)Y_t^{*})+\gamma(P_t-(1-\lambda)P_{t-1})+u_t-(1-\lambda)u_{t-1}=$$

$$=\alpha\lambda+\beta\lambda Y_t+\gamma P_t-(1-\lambda)\gamma P_{t-1}+u_t-(1-\lambda)u_{t-1}$$

This is the model of the same tpe of restricted ADL(1, 1) as in **(b)**

$$C_t=\alpha\lambda+\beta\lambda Y_t+\gamma P_t-(1-\lambda)\gamma P_{t-1}+(1-\lambda)C_{t-1}+u_t-(1-\lambda)u_{t-1}$$

with the same problems of estimation.

9. On the basis of a data $K_i, L_i, Y_i, \quad i = 1, 2, ..., 58$ collected in a survey of 58 similar enterprises the Cobb-Douglas production function

$$Y_i = \theta_0 K_i^{\theta_1} L_i^{\theta_2} e^{\varepsilon_i} \qquad (1)$$

was estimated. It describes dependence of the output Y_i of main factors: capital K_i and labor L_i (all values are expressed in index form, i.e. as percentages of some value taken as a base for comparative analysis). Estimation produced under assumption of a constant returns to scale, gave the following results: $\hat{\theta}_0 = 3.21, \hat{\theta}_1 = 0{,}63, \hat{\theta}_2 = 0.37$.

(a) Comment on an economically meaningful level, what means that the condition of 'constant returns to scale.' Explain what is the meaning of obtained values, bringing the necessary mathematical justification.

(b) Describe the possible econometric procedures that can be used to obtain estimates of the parameters of the Cobb-Douglas production function under assumption of constant returns to scale. Explain in detail how to test the hypothesis of constant returns to scale. What data is necessary for this test and how it could be performed.

(c) Suppose you have time series data $K_t, L_t, Y_t, \quad t = 1, 2, ..., T$ instead of cross-section data. The production function above does not account explicitly the impact of innovation and technological progress. How this function can be modified to analyze the contribution of this factor? What other disadvantages has Cobb-Douglas function, and what alternatives are available for its econometric analysis? How to use existing data to construct the Cobb- Douglas production function in growth rates form? What connections exist between the production function in growth rates form and between the same function in the traditional form?

Solution: (a) Assumption of constant returns to scale can be written in the form $\theta_1 + \theta_2 = 1$. Under this condition production function is homogeneous function of the degree one. It means that if both factors increase by r% the output also increases by r%:

$$Y_i = \theta_0 K_i^{\theta_1} (1 + r/100)^{\theta_1} L_i^{\theta_2} (1 + r/100)^{\theta_2} e^{\varepsilon_i} = (1 + r/100)^{\theta_1 + \theta_2} \theta_0 K_i^{\theta_1} L_i^{\theta_2} e^{\varepsilon_i} = (1 + r/100) \cdot \theta_0 K_i^{\theta_1} L_i^{\theta_2} e^{\varepsilon_i}.$$

The values θ_1 and θ_2 are correspondingly capital and labor elastisities of output. It can be shown by taking logarithms and then differentiating the expression $Y_i = \theta_0 K_i^{\theta_1} L_i^{\theta_2}$ in the variables K_i and L_i:

$$\ln Y_i = \ln \theta_0 + \theta_1 \ln K_i + \theta_2 \ln L_i$$

$$\frac{dY_i}{Y_i} = \theta_1 \frac{dK_i}{K_i}; \quad \frac{dY_i}{Y_i} = \theta_2 \frac{dL_i}{L_i}$$

$$\theta_1 = \frac{\frac{dY_i}{Y_i} 100\%}{\frac{dK_i}{K_i} 100\%}; \quad \theta_2 = \frac{\frac{dY_i}{Y_i} 100\%}{\frac{dL_i}{L_i} 100\%},$$

what are elasticities by definition.

If constant returns to scale is assumed, production function $Y_i = \theta_0 K_i^{\theta_1} L_i^{\theta_2}$ acquires additional properties:

$$Y_i = \theta_0 K_i^{\theta_1} L_i^{\theta_2} = \theta_0 K_i^{\theta_1} L_i^{1-\theta_1} = \theta_0 \left(\frac{K_i}{L_i} \right)^{\theta_1} L_i, \text{ so } \frac{Y_i}{L_i} = \theta_0 \left(\frac{K_i}{L_i} \right)^{\theta_1}$$

The value of $\frac{Y_i}{L_i}$ is the labor productivity of output, and $\frac{K_i}{L_i}$ is capital/labor ratio. Now θ_1 is still capital elasticity of output but evaluated under assumption of constant returns to scale. Estimated values $\hat{\theta}_1 = 0{,}63, \hat{\theta}_2 = 0.37$ mean that for the industry under consideration increase of capital index by 1 percentage point gives $0{,}63$% increase of output, while increase of labor index by 1 percentage point gives only 0.37 % increase of output.

(b) The general (unrestricted) production function $Y_i = \theta_0 K_i^{\theta_1} L_i^{\theta_2} e^{\varepsilon_i}$ is estimated by using OLS after taking logarithms

$$\ln Y_i = \ln\theta_0 + \theta_1 \ln K_i + \theta_2 \ln L_i + \varepsilon_i$$

To estimate $Y_i = \theta_0 K_i^{\theta_1} L_i^{\theta_2} e^{\varepsilon_i}$ under restriction $\theta_1 + \theta_2 = 1$ it is convenient to use expression $\frac{Y_i}{L_i} = \theta_0 \left(\frac{K_i}{L_i}\right)^{\theta_1} e^{\varepsilon_i}$. Taking logarithms we get $\ln\left(\frac{Y_i}{L_i}\right) = \ln\theta_0 + \theta_1 \ln\left(\frac{K_i}{L_i}\right) + \varepsilon_i$. This equation can be estimated using OLS.

To test the restriction $\theta_1 + \theta_2 = 1$ we have to run two regressions and evaluate RSS'

$$\ln \hat{Y}_i = \ln\theta_0 + \theta_1 \ln K_i + \theta_2 \ln L_i, \qquad RSS_U \qquad (1)$$

and

$$\ln\widehat{\left(\frac{Y_i}{L_i}\right)} = \ln\theta_0 + \theta_1 \ln\left(\frac{K_i}{L_i}\right), \qquad RSS_R \qquad (2)$$

and then run F-test

$$F = \frac{(RSS_R - RSS_U)/1}{RSS_U/(58-3)}$$

If $H_0: \theta_1 + \theta_2 = 1$ is rejected one should use (1) for further analysis, otherwise choose (2).

The value of $\hat{\theta}_0 = 3.21$ is just scaling factor, it accumulates (constant) influence of all other factors, not connected directly with the capital and labor.

(c) The value of θ_0 includes the influence of innovations, so it is not realistic to assume that θ_0 is constant among different enterprises. It is also not realistic if the time series data is used instead of cross section sample. For example consider data on the time series $K_t, L_t, Y_t, \ \ t = 1, 2, ..., T$, where K_t, L_t, Y_t are average values of K_{it}, L_{it}, Y_{it} also in index form (as a percentage of some 'initial' value related to the certain year). Then the simplest way to take into account the influence of technological progress on the output is to use data not on set $\theta_0 = A \cdot e^{\gamma t}$

One of possible disadvantage is that CD function is characterized by the elasticity of substitution equal to one. It is too much for many real applications. For example, CES function allows to overcome this disadvantage.

If we take logarithms and differentiate the function $Y_t = A \cdot K_t^{\alpha} \cdot L_t^{\beta} \cdot e^{\gamma t}$, the full differential is:

$$dY_t = d(\ln A + \alpha \ln K_t + \beta \ln L_t + \gamma t) = \alpha \cdot dK_t + \beta \cdot dL_t + \gamma \cdot \mathrm{dt}$$

$$\frac{dY_t}{Y_t} = \alpha \cdot \frac{dK_t}{K_t} + \beta \cdot \frac{dL_t}{L_t} + \gamma \cdot \mathrm{dt}, \text{ or } \frac{Y_t' dt}{Y_t} = \alpha \cdot \frac{K_t' dt}{K_t} + \beta \cdot \frac{L_t' dt}{L_t} + \gamma \cdot \mathrm{dt}$$

and after dividing both parts by dt we get $\frac{Y_t'}{Y_t} = \alpha \cdot \frac{K_t'}{K_t} + \beta \cdot \frac{L_t'}{L_t} + \gamma$. The ratios $\frac{Y_t'}{Y_t}$, $\frac{K_t'}{K_t}$, $\frac{L_t'}{L_t}$ are the continuous growth rates of output, capital and labor Y_t, K_t, L_t respectively. When we substitute the differentials dY_t, dK_t, dL_t (main linear parts of increments) by the increments ΔY_t, ΔK_t, ΔL_t, then we get an approximate formula for discrete rates of growth $y_t = \alpha k_t + \beta l_t + \gamma$, where y_t, k_t, l_t are the discrete growth rates of output, capital and labor. This is the Cobb-Douglas PF in a rate form.

The formulas $Y_t = A \cdot K_t^{\alpha} \cdot L_t^{\beta} \cdot e^{\gamma t}$ and $y_t = \alpha k_t + \beta l_t + \gamma$ are equivalent for continuous time. But the statistical data being used for estimating PF are always discrete. In discrete case the formulas of volume and rate CD functions are not equivalent; they describe different PFs. Even when they are estimated using the same data (i.e. for the growth rates corresponding to the volumes) the results of such estimation may be different.

10. The time series under consideration is given by

$$Y_t = Y_{t-1} - \lambda Y_{t-2} + u_t, \quad 0 < \lambda < 1 \qquad (1),$$

where u_t has zero mean $Eu_t = 0$, constant variance $\operatorname{var} u_t = \sigma^2$ and not correlated with u_{t-k} and Y_{t-k}, $k > 0$.

(a) Use difference Dickey-Fuller equation to show that the series (1) is likely stationary.

(b) Consider now together with the series (1)

$$Y_t = Y_{t-1} - \lambda Y_{t-2} + u_t, \quad 0 < \lambda < 1 \qquad (1),$$

two additional series

$$X_t = X_{t-1} + v_t \qquad (2),$$

and

$$Z_t = -\lambda Z_{t-2} + w_t \qquad (3),$$

where λ in (3) is the same as in (1), v_t and w_t posess the same properties as u_t, that is they have zero expectation, constant variance and they are not correlated with their past values. Additionally we will suppose that u_t, v_t and w_t are distributed independently.
Which of three time series (1), (2) and (3) are cointegrated and which are not?

(c) Find autocorrelation function of Y_t defined in (1) and draw the sketch of its graph for the case $\lambda = 0.5$.
HINT: Multiply equation (1) by Y_{t-1} and take expectations using properties of weak stationarity. Repeat this taking more lags.

Solution: (a) The series $Y_t = Y_{t-1} - \lambda Y_{t-2} + u_t, \quad 0 < \lambda < 1$, can be considered as a special case of the model $Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + u_t$, where $\beta_1 = 1$ and $\beta_2 = -\lambda$.
To apply Dickey-Fuller approach to this series one should rearrange terms like this

$Y_t - Y_{t-1} = \beta_1 Y_{t-1} - Y_{t-1} + \beta_2 Y_{t-2} + u_t$,

$Y_t - Y_{t-1} = \beta_1 Y_{t-1} - Y_{t-1} + \beta_2 Y_{t-1} - \beta_2 Y_{t-1} + \beta_2 Y_{t-2} + u_t$

and

$\Delta Y_t = (\beta_1 + \beta_2 - 1) Y_{t-1} - \beta_2 \Delta Y_{t-1} + u_t$

Now substitute $\beta_1 = 1$ and $\beta_2 = -\lambda$

$\Delta Y_t = (1 - \lambda - 1) Y_{t-1} - \beta_2 \Delta Y_{t-1} + u_t = -\lambda Y_{t-1} + \lambda \Delta Y_{t-1} + u_t$

As $0 < \lambda < 1$ the time series will be classified as stationary (necessary condition of stationarity holds).
Note: in fact this time series is asimptotically stationary, but for simplicity we will say further that it is approximately stationary.

(b) Time series (2) is a random walk, it is nonstationary, but its differences are stationary, so it is integrated of the first order (we can say that it belongs to the class I(1)). The series (1) is stationary as it is shown in **(a)**. The series (3) is also stationary being a special case of AR(2). So their linear combination is also stationary, it means that they are cointegrated. (1) and (3) both belong to the class I(0), the series belonging to different class cannot be cointegrated by definition, so two pairs ((1); (2)) and ((2); (3)) are not cointegrated time series.

(c) The autocorrelation function for the series Y_t is the correlation coefficient $\rho_k = \rho(Y_t, Y_{t+k})$, evaluated for different values of $k > 1$: $\rho_k = \dfrac{\operatorname{cov}(Y_t, Y_{t-k})}{\sqrt{\operatorname{var}(Y_t)\operatorname{var}(Y_{t-k})}}$.

From **(a)** the time series is stationary, so $\operatorname{var}(Y_t) = \operatorname{var}(Y_{t+k})$ and so $\rho_k = \dfrac{\operatorname{cov}(Y_t, Y_{t-k})}{\operatorname{var}(Y_t)}$. We have also from stationarity $\operatorname{cov}(Y_t, Y_{t-1}) = \operatorname{cov}(Y_{t-1}, Y_{t-2}) = \operatorname{cov}(Y_{t-2}, Y_{t-3}) = \ldots$, and $\operatorname{cov}(Y_t, Y_{t-2}) = \operatorname{cov}(Y_{t-1}, Y_{t-3}) = \ldots$, and so

on (covariance depends only on the length of the lag but not on the time). As also $E(Y_t) = E(Y_{t-1}) = E(Y_{t-2})$ then from $Y_t = Y_{t-1} - \lambda Y_{t-2} + u_t$ and $Eu_t = 0$ follows that $E(Y_t) = 0$.
Multiplying both sides of the equation (1)

$$Y_t = Y_{t-1} - \lambda Y_{t-2} + u_t \qquad (1)$$

by Y_{t-1} and taking expectations
we get

$$E(Y_t, Y_{t-1}) = E(Y_{t-1}^2) - \lambda E(Y_{t-1}, Y_{t-2}) + E(u_t, Y_{t-1})$$

or

$$\text{cov}(Y_t, Y_{t-1}) = \text{var}(Y_{t-1}) - \lambda \,\text{cov}(Y_{t-1}, Y_{t-2})$$

as $\text{cov}(Y_{t-k}, u_{t-s}) = 0,\ s \neq k$.
Changing indices in the last two terms and dividing both sides of the equation by $\text{var}(Y_t)$,

$$\frac{\text{cov}(Y_t, Y_{t-1})}{\text{var}(Y_t)} = 1 - \lambda \frac{\text{cov}(Y_t, Y_{t-1})}{\text{var}(Y_t)}$$

or

$$\rho_1 = 1 - \lambda \rho_1$$

and so

$$\rho_1 = \frac{1}{1+\lambda}.$$

Repeating this procedure with multiplying (1) by Y_{t-2} we get

$$\text{cov}(Y_t, Y_{t-2}) = \text{cov}(Y_{t-1}, Y_{t-2}) - \lambda \,\text{var}(Y_{t-2})$$

$$\frac{\text{cov}(Y_t, Y_{t-2})}{\text{var}(Y_t)} = \frac{\text{cov}(Y_t, Y_{t-1})}{\text{var}(Y_t)} - \lambda$$

or

$$\rho_2 = \rho_1 - \lambda$$

and so

$$\rho_2 = \frac{1}{1+\lambda} - \lambda = \frac{1-\lambda-\lambda^2}{1+\lambda}$$

Repeating this again now for Y_{t-k} one can find recurrence

$$\text{cov}(Y_t, Y_{t-k}) = \text{cov}(Y_{t-1}, Y_{t-(k-1)}) - \lambda \,\text{var}(Y_{t-(k-2)})$$

and so

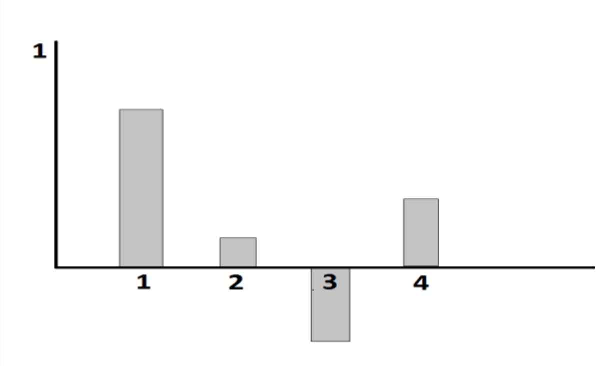
$$\rho_k = \rho_{k-1} - \lambda \rho_{k-2}$$

From this recurrence all autocorrelation coefficients can be found, for example,

$$\rho_3 = \rho_2 - \lambda\rho_1 = \frac{1-\lambda-\lambda^2}{1+\lambda} - \lambda\frac{1}{1+\lambda} = \frac{1-\lambda-\lambda^2-\lambda-\lambda^2}{1+\lambda} = \frac{1-2\lambda-2\lambda^2}{1+\lambda}$$

$$\rho_4 = \rho_3 - \lambda\rho_2 = \frac{1-2\lambda-2\lambda^2}{1+\lambda} - \lambda\frac{1-\lambda-\lambda^2}{1+\lambda} = \frac{1-2\lambda-2\lambda^2-\lambda+\lambda^2+\lambda^3}{1+\lambda} = \frac{1-3\lambda-\lambda^2+\lambda^3}{1+\lambda}$$

For the case $\lambda = 0.5$ $\rho_1 = \dfrac{2}{3},\ \rho_2 = \dfrac{1}{6},\ \rho_3 = \dfrac{-1}{3},\ \rho_4 = \dfrac{1}{3}, \ldots$



Sketch of the graph of autocorrelation function

The International College of Economics and Finance
Econometrics – 2014-2015.
Midterm exam 2015. March 26.

General instructions. **Candidates should answer EIGHT of the following TEN questions: all questions of the Section A and any three of the questions from Section B (questions 6-10). The weight of the Section A is 40% of the exam; three other questions from the Section B add 20% each. You are advised to divide your time accordingly. Structure your answers in accordance with the structure of the questions. When testing hypotheses always state clearly null and alternative hypotheses provide critical value used for test, mentioning degrees of freedom and the significance level chosen for the test.**

SECTION A

Answer **ALL** questions from this section (questions **1-5**).

1. A researcher has data on annual household expenditure on home video, Y, relative price index on home video RPI, annual household wage income, W, and annual household non-wage income, NW, all measured in thousands of U.S. dollars (except price index, that is measured as index numbers equal to 100 in the year 2000), for the period 1990-2008. Problems associated with non-stationary time series may be ignored.

He is considering fitting the model

$$Y_t = \alpha + \beta_1 W_t + \beta_2 NW_t + \beta_3 RPI_i + u_t$$

The correlation between W and NW is 0.97.

Estimation of the equation gives the following result (standard errors in parentheses; RSS is residual sum of squares):

$$\underset{}{\hat{Y}_t} = \underset{(5.21)}{-12.48} + \underset{(0.025)}{0.06} \cdot W_t + \underset{(0.02158)}{0.04439} \cdot NW_t - \underset{(0.77)}{3.28} \cdot RPI_t \qquad R^2 = 0.98 \qquad (1)$$

Comment the estimated equation, paying attention to the interpretation of the coefficients, significance of the coefficients and equation as a whole, and potential problems.

Solution and marking

2 marks for correct interpretation of coefficients. As wage income rises by 1 thousand US dollars the expenditure on home video rises on average by 60 US dollars. Analogously for non-wage income. As relative price index rises by 1 percentage point the expenditure on home video falls on average by 3,280 US dollars.

2 mark for evaluation of t-statistics: t-statistics for W coefficient is 2.4, that is significant for df = 19 - 4 = 15 only at 5% significance level ($t_{crit}^{5\%}(df=15) = 2.131$, $t_{crit}^{1\%}(df=15) = 2.947$). t-statistics for NW coefficient is 2.057 that is insignificant. Coefficient of RPI is significant at any reasonable significance level

3 mark for the discussing of R-squared and evaluation of F-statistics on the base of R-squared – $F = \dfrac{0.98/3}{(1-0.98)/15} = 245$ that is significant at 1% significance level ($F_{crit}^{1\%}(3,15) = 5.42$).

1 mark for discussing the potential problem of multicollinearity (it was mentioned that the correlation between correlation between W and NW is 0.95). High correlation between explanatory variables not always leads to some manifestation of the multicollinearity. Here it affects significantly the quality of the regression possibly making one of the coefficients insignificant and the other – weakly significant).

[Total 8 marks].

2. Researcher estimates two regression models

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + u_t \text{ and}$$
$$Y_t = \beta_1' + \beta_2' X_t + \beta_3' Y_{t-1} + v_t,$$

The variable X_t is supposed to be non-stochastic. But she suspects that disturbance terms of these models are subject first order autocorrelation process: $u_t = \rho u_{t-1} + \varepsilon_t$, $v_t = \rho' v_{t-1} + \varepsilon_t$ where $|\rho| < 1$, $|\rho'| < 1$, $E(\varepsilon_t) = 0$; $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ and $E(\varepsilon_s \varepsilon_t) = 0$ for all *s* and *t*. Compare the consequences of using ordinary least square for estimation both equations. How to detect autocorrelation here?

Solution and marking

If error terms of both regression models satisfy Gauss-Markov conditions OLS would give BLUE estimates. If the error term u_t in the regression model $Y_t = \beta_0 + \beta_1 X_t + ... + u_t$; $t = 1, 2, ..., T$ is such that $E(u_t u_s) \neq 0$, then the error term is said to be serially correlated. This a very general condition and it is generally necessary to assume the more restrictive condition that $u_t = \rho u_{t-1} + \varepsilon_t$ where $|\rho| < 1$ and $E(\varepsilon_t) = 0$; $E(\varepsilon_t^2) = \sigma_\varepsilon^2$ and $E(\varepsilon_s \varepsilon_t) = 0$ for all *s* and *t*. Usual consequences of serial correlation for the first equation is that the ordinary least squares (OLS) parameter estimates are unbiased and consistent but inefficient and their standard errors (and hence *t* and *F*-values) are incorrect. But if there is a lagged dependent variable as a explanatory variable in the model $Y_t = \beta_1' + \beta_2' X_t + \beta_3' Y_{t-1} + v_t$ OLS parameter estimates are inconsistent. In the presence of lagged dependent variable usual Durbin-Watson statistics becomes biased to the value equal 2, so in this situation Durbin *h*-statistics should be applied instead: $h = \hat{\rho} \cdot \sqrt{\frac{n}{1 - n \cdot S_{\beta_3}{}^2}}$. Here $\hat{\rho}$ is the supposed autocorrelation coefficient (in equation $u_t = \rho u_{t-1} + \varepsilon_t$) and could be approximately estimated using usual Durbin-Watson d-statistics: $\hat{\rho} = (1 - 0.5d)$. $S_{\beta_3}{}^2$ in the formula for *h*-statistics means the square of standard error for coefficient of lagged dependent variable Y_{t-1}, and number of observations is evaluated taking into account that using lagged variables diminiches total number of observation (by one in our case). Under assumption of no serial correlation *h*-statistics follows the standard normal *z*-distribution, so usual critical value could be used (1.96 for significance level 0.05, for example). Quite a rare problem with this statistics is that expression under square root $1 - n \cdot S_{\beta_3}{}^2$ could be negative (usually in small samples), then usual d-statistics could be studied with some reservations. **[Total 8 marks].**

3. A researcher has data on the average annual rate of growth of employment, *e*, and the average annual rate of growth of GDP, *x*, both measured as percentages, for a sample of 17 developing countries and 13 developed ones for the period 1995-2008. She runs first simple regressions of *e* on *x*

for the whole sample $\quad \hat{e}_i = -0{,}56 + 0.24 x_i \quad R^2 = 0.17$

$\qquad\qquad\qquad\qquad (0.53)\ (0.10)$

Then she defines a dummy variable *D* which is equal to 1 for the developing countries and 0 for the others. Hypothesizing that the impact of GDP growth on employment growth is lower in the developed countries than in the developing ones, she defines a slope dummy variable *xD* as the product of *x* and *D* and fits the regression (standard errors in parentheses):

$$\hat{e}_i = -1{,}43 + 0.19 x_i + 0.52 D_i + 0.78 (xD)_i \quad R^2 = 0.36$$
$$(0.36)\ (0.07) \quad (0.33) \quad (0.42)$$

Is the difference between developing and developed countries significant in evaluating the influence of the average annual rate of growth of GDP on the average annual rate of growth of employment?
Was the researcher right in her hypothesis on the impact of GDP growth?

Solution and marking

The dummy variables allow the intercept and the slope coefficient to be different for developing and developed countries. [**1 marks].** To evaluate the significance of these difference one could use F-test for a group of dummies. Comparing R-squared for two regression we get:

$$F(2,46) = \frac{(0.36 - 0.17)/2)}{(1 - 0.36/46} = 3.86$$

The critical value of F(2,26) at the 5% significance level is 3.37. Hence the null hypothesis that the coefficients are the same for developed and developing countries is rejected. **[4 marks]**

The second part of the question refers to the one-sided test. As F-test, based on the sum of squares cannot evaluate one-sided effects, one could use one-sided t-tests for both dummies. As $t_{crit}^{5\%}(one\ sided, 26) = 1.706$ coefficient of shift dummy is insignificant ($t = \frac{0.52}{0.33} = 1.576$) while the slope dummy is significant at 5% ($t = \frac{0.78}{0.42} = 1.86$). [**3 marks].**

[Total 8 marks].

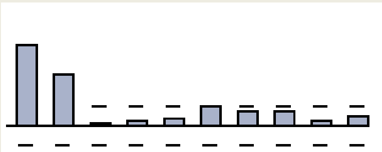
.

4. Discuss the term "stationary time series". Show that MA(2) process $X_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2}$ is stationary. Why is it important to know whether the time series used in the regression model are stationary? What is autocorrelation function? How to recognize MA(2) process looking on the graphical representation of its autocorrelation function? Explain.

[8 marks] According to the definition, the time series X_t is said to be weakly stationary if its expected value and population variance are independent of time and if the population covariance between its values at time t and $t+r$ depends on r but not on time. As $EX_{t+r} = 0$ for any r $EX_t = E\varepsilon_t + \alpha_1 E\varepsilon_{t-1} + \alpha_2 E\varepsilon_{t-2} = 0$. If it is known that $\operatorname{Var} X_{t+r} = \sigma^2$ and as $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}$ are uncorrelated $\operatorname{Var} X_t = \operatorname{Var}(\varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2}) = \alpha_1^2 \operatorname{Var} \varepsilon_{t-1} + \alpha_2^2 \operatorname{Var} \varepsilon_{t-2} = \alpha_1^2 \sigma^2 + \alpha_2^2 \sigma^2$ does not depend on time. Next the population covariance between X_t and X_{t-r} is zero for all $r > 2$ since then X_t and X_{t-r} have no elements in common. For $r = 1$ and $r = 1$ as $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \varepsilon_{t-4}$ are uncorrelated both expressions $\operatorname{Cov}(X_t, X_{t-1}) = \operatorname{Cov}(\varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2}, \varepsilon_{t\text{-}1} + \alpha_1 \varepsilon_{t-2} + \alpha_2 \varepsilon_{t-3}) = \alpha_1 \sigma^2 + \alpha_1 \alpha_2 \sigma^2$ and $\operatorname{Cov}(X_t, X_{t-2}) = \operatorname{Cov}(\varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2}, \varepsilon_{t\text{-}2} + \alpha_1 \varepsilon_{t-3} + \alpha_2 \varepsilon_{t-4}) = \alpha_2 \sigma^2$ does not depend on time. In fact we know that all MA processes are stationary.

The autocorrelation function for the series Y_t is the correlation coefficient $\rho_k = \rho(Y_t, Y_{t+k})$, evaluated for different values of $k > 1$: $\rho_k = \frac{\operatorname{cov}(Y_t, Y_{t-k})}{\sqrt{\operatorname{var}(Y_t)\operatorname{var}(Y_{t-k})}}$.

Typical correllogram looks like this having just two positive 'teeth' and then falling abruptly to zero.



[Total 8 marks].

5. The student has data comprising 6 developed countries for the period of 50 consecutive years: expenditures on travel abroad per capita E_{it}, income per capita Y_{it} and relative prices for travel abroad P_{it}. This data constitute a panel. The model under investigation be

$$\log E_{it} = \beta_1 + \beta_2 \log(Y_{it}) + \beta_3 \log(P_{it}) + \alpha_i + u_{it};\ i = 1, 2, ..., 6;\ t = 1, 2, ..., 50$$

The student runs three alternative approaches to the evaluation of this model: 1) pooled OLS regression, 2) fixed effects panel regression model, and 3) random effects panel regression model.

(a) Help the student to choose between different approaches: for each of three pairs of alternatives indicate corresponding test, null hypothesis, type of distribution of the test statistic, the number of degrees of freedom and the decision rule (which alternative is chosen if null hypothesis is rejected and which if it is not). What are advantages and risks of each choice?

(b) The model above is based on the assumption that the income and price elasticities for different countries are the same. Suggest how to test this assumption if in fact there is no unobserved heterogeneity. Give some details.

Solution: (a) Of course, only inexperienced student could consider three different types of models together. In practice there is always a certain logic in the sequence of the models and tests applied.
First the student is recommended to use Darbin-Wu-Hausman (DWH) test to choose between fixed and random effects. It is standard for majority of econometric computer programs and is based on using chi-square statistics with degrees of freedom equal to the number of variables in the equation under consideration (2 in our case) (as it compares estimates of coefficients obtained by two alternative models). Under H_0 that there is no difference between coefficients obtained by two alternative models – fixed and random panel models (which means that unobserved heterogeneity α_i as a part of disturbance term, is not correlated with $\log(Y_{it})$ and $\log(P_{it})$) both fixed effect and random effect models provide us with consistent estimates. We choose in this case random effect models as it retains in disturbance term all unobserved heterogeneity, there is no reduction of degrees of freedom typical for fixed effects models.
If H_0 is rejected, so there are essential differences between coefficients obtained using fixed and random effects models, we choose fixed effects model, because rejecting of H_0 means that main assumption of independence of the disturbance term from regressors is violated so using random effects model we are under risk of getting inconsistent estimates of parameters.
So we have to suffice the fixed effects model that always gives consistent estimates.
Note that fixed effect model drops intercept (all factors that are constant in time) and includes a set of additional dummy variables (LSDV model) so it involves estimating an additional set of coefficients, that leads to reduction of degrees of freedom and so sometimes we observe more insignificant coefficients (for example in first difference or in within group versions of fixed effect models).

Further choice of the pair of model under consideration depends on the result of this test.
If the random effect model is chosen on the base of DWH test, it is possible that ordinary least squares is even better in efficiency if there is in fact no unobserved heterogeneity and so there no random effects at all. There are some tests for this purpose, for example Breush-Pagan test based on Lagrange Multiplier approach. It also uses chi-square distribution with degrees of freedom equal to 1 under H_0 of the absence of random effects.
If fixed effect model is chosen then conventional Chow test could be performed on the base of comparison of pooled regression with 6 separate regressions for $i = 1$, $i = 2$, …, and $i = 6$ under H_0 that there is no unobserved heterogeneity. If H_0 is rejected we choose fixed effect model, and we choose pooled regression if not.

(b) To answer this question is simpler in case of LSDV fixed effects model of the type

$$\log E_{it} = \beta_1 + \beta_2 \log(Y_{it}) + \beta_3 \log(P_{it}) + \sum_{i=1}^{6} \gamma_i D_i + u_{it}. \qquad (1)$$

where D_i are dummies corresponding different goods under consideration. To take into account that the elasticities of DPI_t and $PRGOOD_{it}$ could be different, it possible to run instead 6 different regressions for 6 goods ($i = 1, 2, ..., 6$),

$$\log E_{it} = \beta_1 + \beta_2 \log(Y_{it}) + \beta_3 \log(P_{it}) + u_{it}$$

Then perform an F-test using the values of RSS' of evaluated models:

$$F = \frac{(RSS_{LSDV} - \sum_{i=1}^{6} RSS_i)/(6 \cdot 3 - (2+6))}{\sum_{i=1}^{6} RSS_i /(50 - 6 \cdot 3)} = \frac{(RSS_{LSDV} - \sum_{i=1}^{6} RSS_i)/10}{\sum_{i=1}^{6} RSS_i /32}.$$

Here RSS_{LSDV} - the value of RSS for LSDV fixed effect model, $\sum_{i=1}^{6} RSS_i$ - sum of RSS' for separate regressions for different goods, the rule for degrees of freedom follows general principles of F-test and is clear from the formula above. If H_0 is not rejected there is no significant differences between elasticities. **[Total 8 marks].**

SECTION B

Answer **THREE** questions from this section (questions **6-10**).

6. The researcher runs two production function models for the same data for some developing country: $t = 1,2,...30$, where y_t is income per capita, x_t is a capital, and z_t is labor (all variables are index numbers)

$$\ln y_t = \alpha + \beta \ln x_t + 0.5 \ln z_t + u_t \qquad (1)$$

$$\ln y_t = \alpha + \beta \ln x_t + \beta \ln z_t + v_t \qquad (2)$$

given that x_t and z_t are deterministic sequences and $u_t \sim iid(0,\sigma^2)$, $v_t \sim iid(0,\sigma^2)$.

(a) Explain how to find the least squares estimates of β. What are estimators of β for both equations (write out the explicit formulas)?

a) Both regressions are in fact simple linear regression models, so in both cases one should use conventional estimator of the type $\hat{\beta} = \dfrac{\text{Cov}(Y, X)}{\text{Var}(X)}$. But the transformations of the data needed to reduce equation to the simple regression model, are different, hence estimators will be different.
For (1) one can rewrite the model in the form $\ln y_t - 0.5 \ln z_t = \alpha + \beta \ln x_t + u_t$ and then estimate the model $y_t^{\cdot} = \alpha + \beta \ln x_t + u_t$, where $y_t^{\cdot} = \ln y_t - 0.5 \ln z_t$.
The second equation could be estimated after simple transformation $\ln y_t = \alpha + \beta(\ln x_t + \ln z_t) + v_t$, so the model is $\ln y_t = \alpha + \beta x_t^{\cdot} + u_t$ where $x_t^{\cdot} = \ln x_t + \ln z_t$.
It means that OLS gives best estimators:
for (1) we get $\hat{\beta} = \dfrac{\text{Cov}(\ln y_t - 0.5 \ln z_t, \ln x_t)}{\text{Var}(\ln x_t)} = \dfrac{\text{Cov}(\ln y_t, \ln x_t) - 0.5\,\text{Cov}(\ln z_t, \ln x_t)}{\text{Var}(\ln x_t)}$
for (2) we get $\hat{\beta} = \dfrac{\text{Cov}(\ln y_t, \ln x_t + \ln z_t)}{\text{Var}(\ln x_t + \ln z_t)} = \dfrac{\text{Cov}(\ln y_t, \ln x_t) + \text{Cov}(\ln y_t, \ln z_t)}{\text{Var}(\ln x_t) + \text{Var}(\ln z_t) + 2\,\text{Cov}(\ln x_t, \ln z_t)}$. **[5 marks]**

(b) Both regressions are the restricted versions of the general model

$$\ln y_t = \alpha + \beta \ln x_t + \gamma \ln z_t + u_t \qquad (3).$$

What are the restrictions? How they could be tested? What are possible outcomes of two tests?

b) As it is said in the text both models are the restricted versions of the general model

$$y_t = \alpha + \beta x_t + \gamma z_t + u_t$$

The restrictions for the models are correspondingly $\gamma = 0.5$ and $\gamma = \beta$. One first should test each restriction using F-test on the base of RSS or R^2 of each equation. For example, to test the restriction $\gamma = \beta$ we use F-statistics $F = \dfrac{(RSS_2 - RSS_3)/1}{RSS_3/(30-3)}$ where RSS_2 and RSS_3 are the sums of squared residuals for the models (2) and (3). The same for the restriction $\gamma = 0.5$ (use RSS_1 and RSS_3). Note that the restriction $\gamma = 0.5$ also could be testet using t-test with t-statistic $e = \dfrac{g - 0.5}{s.e.(g)}$ where g is OLS estimator of γ in the equation (3) (test uses $df = 30 - 3$). Possible outcomes of these tests are: one should estimate the general model if both restrictions are invalid, the model with valid restriction if only one is valid (which is superior for this case). If both restrictions are valid, any of them can be estimated, but the one with smaller variance is superior. **[5 marks]**

(c) What are means and variances of the estimators in (a)?

c) If both restrictions are valid the estimator of β is unbiased, and it is biased if the restriction leading to the model is invalid.
The variance for the estimator of regression coefficient in the model $y_t = \beta_1 + \beta_2 x_t + u_t$ is given by expression $\sigma^2_{\hat{\beta}_2} = \dfrac{\sigma_u^2}{n\,\mathrm{Var}(X)} = \dfrac{\sigma^2}{n\,\mathrm{Var}(X)}$. For the model (1) (if it is valid) $\sigma^2_{\hat{\beta}_2} = \dfrac{\sigma_u^2}{n\,\mathrm{Var}(\ln x_t)}$; for the model (2) $\sigma^2_{\hat{\beta}_2} = \dfrac{\sigma_v^2}{n\,\mathrm{Var}(\ln x_t + \ln z_t)} = \dfrac{\sigma^2}{n\,\mathrm{Var}(\ln x_t + \ln z_t)}$ (as the variances of the error terms are supposed to be equal). **[5 marks]**

(d) Suppose that both restrictions in (b) are not rejected. Under what conditions the estimator of β from equation (1) is superior to that from equation (2) (and vice versa) using the criteria of minimum variance of the estimator? Explore different cases:

i) $\ln x_t$ and $\ln z_t$ are not correlated (or nearly not correlated)

ii) $\ln x_t$ and $\ln z_t$ are positively correlated (perfectly or not)

iii) $\ln x_t$ and $\ln z_t$ are negatively correlated (perfectly or not)

d) If both models are valid, the choice of the model would depend on the comparison of two expressions for the variances $\mathrm{Var}(\ln x_t)$ and $\mathrm{Var}(\ln x_t + \ln z_t)$.
For example if $\ln x_t$ and $\ln z_t$ are not correlated (or nearly not correlated) $\mathrm{Var}(\ln x_t + \ln z_t) = \mathrm{Var}(\ln x_t) + \mathrm{Var}(\ln z_t) > \mathrm{Var}(\ln x_t)$ so the regression (2) is superior. The same is true for the case of positive correlation $\mathrm{Var}(\ln x_t + \ln z_t) = \mathrm{Var}(\ln x_t) + \mathrm{Var}(\ln z_t) + \mathrm{Cov}(\ln x_t, \ln z_t) > \mathrm{Var}(\ln x_t)$. For the case of negative correlation the result is uncertain as $\mathrm{Var}(\ln x_t + \ln z_t) = \mathrm{Var}(\ln x_t) + \mathrm{Var}(\ln z_t) - \mathrm{Cov}(\ln x_t, \ln z_t)$, but one could be supposed that in some cases $\mathrm{Var}(\ln x_t + \ln z_t) < \mathrm{Var}(\ln x_t)$ so the estimator based on the regression (1) is superior. **[5 marks]**
[Total 20 marks]

7. The researcher tries to choose between two alternative models, explaining the dynamics of the total expenditure on jewelry J_t in USA for 1983-2007 and the influence of the prices P_t (nominal price index) on the demand on jewelry. First explanation is based on the adaptive expectation model

$$J_t = \beta_1 + \beta_2 P_{t+1}^e + u_t \qquad \text{(AE1)},$$

where P_{t+1}^e stands for expectations of the value of P for the future period $t+1$, the dynamics of that is described by the following adaptive expectations equation

$$P_{t+1}^e - P_t^e = \lambda(P_t - P_t^e) \qquad \text{(AE2)}.$$

Second explanation uses partial adjustment model

$$J_t^* = \beta_1 + \beta_2 P_t + u_t \qquad \text{(PA1)}.$$

where J_t^* represents 'target' or 'desired' level of expenditure on jewlery, while actual level J_t is subject the adjustment process

$$J_t - J_{t-1} = \lambda(J_t^* - J_{t-1}) \qquad \text{(PA2)}.$$

Then she added dependent lagged variable

$$\hat{J}_t = \underset{(0.18)}{0.36} - \underset{(0.003)}{0.003}P_t + \underset{(0.060)}{1.052}J_{t-1} \qquad R^2 = 0.980, \quad DW = 1.81 \quad RSS = 1.27 \qquad (2)$$

After additionally added lagged prices the equation looks as

$$\hat{J}_t = \underset{(0.159)}{0.204} - \underset{(0.007)}{0.023}P_t + \underset{(0.006)}{0.02}P_{t-1} + \underset{(0.051)}{1.083}J_{t-1} \qquad R^2 = 0.987, \quad DW = 2.19 \quad RSS = 0.842 \qquad (3)$$

Being upset by the wrong sign of the coefficient for lagged prices she ran also restricted version of the same model

$$\hat{J}_t = \underset{(0.191)}{0.308} - \underset{(0.001)}{0.0009}P_t + 0.0009 \cdot 1.022 P_{t-1} + \underset{(0.059)}{1.022}J_{t-1} \qquad R^2 = 0.979, \quad DW = 1.75 \quad RSS = 1.332 \qquad (4)$$

(a) How the partial adjustment model (PA1-PA2) could be evaluated on the basis of the estimation results. Describe briefly without excessive math. What are the properties of estimated coefficient?

Solution and marking

a) The partial adjustment model can be evaluated on the basis of regression (2), where the term with lagged dependent variable is present. In principle if to substitute the regression (PA1) $J_t^* = \beta_1 + \beta_2 P_t + u_t$ into equation of adjustment (PA2) $J_t - J_{t-1} = \lambda(J_t^* - J_{t-1})$, one can get $J_t - J_{t-1} = \lambda(\beta_1 + \beta_2 P_t + u_t - J_{t-1})$ and so $J_t = \lambda\beta_1 + \lambda\beta_2 P_t + \lambda u_t + (1-\lambda)J_{t-1} + \lambda u_t$. Considering (2) as partial adjustment model one can evaluate $1-\lambda = 1.052$, so $\lambda = -0.052$ then $\lambda\beta_2 = (-0.052)\cdot\beta_2 = -0.003$ so $\beta_2 = 0.058$. By analogy $\lambda\beta_1 = (-0.052)\cdot\beta_1 = 0.36$ then $\beta_1 = -6.92$ (*it was not expected that students to present this math, the general idea would be enough*).

It is essentially to note that if the disturbance term is not subject autocorrelation OLS estimators of the regression equation in the partial adjustment model would be consistent (the disturbance term of equation (PA1) is merely the same as disturbance term in (2)). It could be mentioned that although both slope coefficients of the regression equation (2) have correct sigh, the coefficient of prices is insignificant, so some other models should be considered.

[Total 5 marks for a].

(b) How the adaptive expectation model (AE1-AE2) could be evaluated on the basis of the estimation results. Some comments will be enough. What are the properties of estimated coefficient?

Solution and marking

b) The adaptive expectation model is also fitted by equation (2), now using Koyck transformation:
First step – substitution of expectation mechanism into regression

$$P_{t+1}^e - P_t^e = \lambda(P_t - P_t^e) \Rightarrow P_{t+1}^e = P_t^e + \lambda P_t - \lambda P_t^e = \lambda P_t + (1-\lambda)P_t^e$$

$$J_t = \beta_1 + \beta_2 P_{t+1}^e + u_t = \beta_1 + \beta_2(\lambda P_t + (1-\lambda)P_t^e) + u_t$$

$$J_t = \beta_1 + \beta_2 \lambda P_t + (1-\lambda)\beta_2 P_t^e + u_t \qquad (*)$$

Second step –lagging regression

$$J_{t-1} = \beta_1 + \beta_2 P_t^e + u_{t-1} \Rightarrow \beta_2 P_t^e = \beta_1 - J_{t-1} + u_{t-1}$$

and substituting it into transformed (*)

$$J_t = \beta_1 + \beta_2 \lambda P_t + (1-\lambda)(\beta_1 - J_{t-1} + u_t) + u_t$$

$J_t = \beta_1 + (1-\lambda)\cdot\beta_1 + \beta_2\lambda P_t - (1-\lambda)J_{t-1} + u_t + (1-\lambda)\cdot u_{t-1}$ (*this math deserves bonus mark*).

But now the disturbance term after Koyck transformation is subject the autocorrelation (the disturbance term follows autocorrelation process of the moving average type). So the OLS estimates of the regression (2) and would be inconsistent even if u_t is not subject autocorrelation.

[Total 5 marks for b].

(c) The researcher's wants to choose between partial adjustment model and adaptive expectation model. He plans to use for this purpose the information on the coefficients of estimated regressions and also information whether their disturbance terms are under autocorrelation or not. Comment on the researcher's plan explaining what is true and what could be wrong in it.

Solution and marking

c) The idea to use coefficients (of the model (2) of course) is totally wrong, as after some transformations both models (partial adjustment and adaptive expectations) gives the same in its structure) regression model. The models (3) and (4) has nothing to do with either models, they are simply ADL(1, 1) models and useless for the desired purpose. As for autocorrelation, the idea has some sense, as we know that Koyck transformation applied to the adaptive expectations model leads to autocorrelation. But it is possible that initial disturbance term of the equation (AE1) or (PA1) is subject autocorrelation, so that transformed equation (2) would inherits this property. It means that if the autocorrelation in equation (2) is detected it could be equally likely be the attributed to the Koyck transformation (based on the assumption of (AE) model), and as well to the fact that equation (PA1) suffers from autocorrelation (based on the assumption of (PA) model).

On the other hand it is possible that in the situation of autocorrelated disturbance term in (AE1) $u_t = \rho u_{t-1} + \varepsilon_t$ the influence of ρ would be balanced by other component in a composite disturbance term after Koyck transformation $u_t - (1-\lambda)u_{t-1} = \rho u_{t-1} + \varepsilon_t - (1-\lambda)u_{t-1} = \varepsilon_t + (\rho + \lambda - 1)u_{t-1} \approx \varepsilon_t$ if $\rho + \lambda \approx 1$ simply by chance. It means that if the autocorrelation in equation (2) is not detected it could be assigned to the fact that it is based on assumptions of PA model with non-correlated disturbance term in (PA1), but it could be also assigned to the balanced AE model with autocorrelated disturbance term in (AE1).

In fact there is probably no autocorrelation as Durbin statistics $d = \hat{\rho}\sqrt{\dfrac{n}{1 - ns_{Y_{t-1}}^2}}$ ($\hat{\rho}$ is an estimate of the autocorrelation coefficient, usually $\hat{\rho} = 1 - 0.5 \cdot DW$), $s_{Y_{t-1}}^2$ is the standard deviation of the coefficient estimate of lagged dependent variable, n is the corrected number of observations) for the equation (2) gives $d = (1 - 0.5*1.81)\cdot\sqrt{\dfrac{24}{1 - 24*0.06^2}} = 0.48$ what is less than $z_{crit}^{5\%} = 1.96$.

[Total 5 marks for c].

(d) Perform common factor test and make your decision. What is interpretation of the coefficient for lagged dependent variable? What is final conclusion from the whole analysis?

Solution and marking

d) The common factor test relates to the ADL(1, 1) models (3) and (4),

$$J_t = \beta_1 + \beta_2 P_t + \beta_3 P_{t-1} + \beta_4 J_{t-1} + u_t \qquad (3^*)$$

$$J_t = \beta_1 + \beta_2 P_t - \beta_2 \cdot \beta_4 P_{t-1} + \beta_4 J_{t-1} + u_t \qquad (3^*)$$

which has the same set of variables but the model (4) uses the restriction $\beta_3 = -\beta_2 \cdot \beta_4$.

The test statistics is $\chi^2 = n \cdot \ln(RSS_{restricted} / RSS_{unrestricted})$, where n should be adjusted according to the lag structure. Using data provided we get $\chi^2 = 24 \cdot \ln(1.332/0.842) = 11.01$. As chi-square critical at 0.1% significance (here df=1, as the number of restrictions is 1) is 10.828 the null hypothesis of the validity of the restriction is rejected at any reasonable significance level. So unrestricted version of ADL(1, 1) model (equation (3)) should be chosen for estimation and analysis. The interpretation of the coefficient for lagged dependent variable is something like influence of the former consumption of jewelry items on current consumption. We would have interpreted this coefficient as autocorrelation coefficient if the restriction were valid so we could have assumed the autocorrelated disturbance term.

[Total 5 marks for d].

8. The student having data (daily time series) on the official exchange rate (ruble per dollar) RUR_USD and the price of oil (Brent) OIL_BRENT (in dollars per barrel) from January 01/2014 to March 15/2015, tries to understand their relationship. Suspecting that time series OIL_BRENT is not stationary he runs Dickey-Fuller test for three different periods separately:

Period I: from January 01/2014 to August 29/2014;
Period II: from September 01/2014 to December 31/2014;
Period III: from January 01/2015 to March 15/2014.

a) First he runs Dickey-Fuller test (no additional lags, no trend, no intercept) for the period I and found the following

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-0.335556	0.5630

Explain what is Dickey-Fuller test for this case, derive its equation, state hypotheses, and help the student to interpret obtained results.

Solution: (a) Starting from the equation above $X_t = \beta_1 X_{t-1} + u_t$ we subtract the same term X_{t-1} from both sides:

$$X_t - X_{t-1} = \beta_1 X_{t-1} - X_{t-1} + u_t$$

taking out the common factors of the brackets

$$X_t - X_{t-1} = (\beta_1 - 1) X_{t-1} + u_t$$

or using differences

$$\Delta X_t = (\beta_1 - 1) X_{t-1} + u_t$$

The ADF test is based on the null hypothesis of no stationarity

$H_0 : \beta_1 - 1 = 0$ against $H_a : \beta_1 - 1 < 0$ and uses special ADF statistics, with critical values calculated by computer. If the value of ADF less than critical null hypothesis is rejected.

In our case P-value for ADF statistics is quite large so the null hypothesis of non stationarity can not be rejected. **[5 marks]**

b) He uses also augmented Dickey-Fuller test including one additional lag: $X_t = \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_t$ where u_t is distributed independently of X_t with zero mean and constant variance. Derive Dickey-Fuller equation for this case, state hypotheses and decision rule

Solution: (a) Starting from the equation above $X_t = \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_t$ we first subtract the same term X_{t-1} from both sides:

$$X_t - X_{t-1} = \beta_1 X_{t-1} - X_{t-1} + \beta_2 X_{t-2} + u_t$$

Then add to and simultaneously subtract from the right side of equation the same term $\beta_2 X_{t-1}$:

$$X_t - X_{t-1} = \beta_1 X_{t-1} - X_{t-1} + \beta_2 X_{t-1} - \beta_2 X_{t-1} + \beta_2 X_{t-2} + u_t$$

and rearrange terms taking out the common factors of the brackets

$$X_t - X_{t-1} = (\beta_1 + \beta_2 - 1) X_{t-1} - \beta_2 (X_{t-1} - X_{t-2}) + u_t$$

or using differences

$$\Delta X_t = (\beta_1 + \beta_2 - 1) X_{t-1} - \beta_2 \Delta X_{t-1} + u_t$$

The ADF test is based on the null hypothesis of no stationarity $H_0 : \beta_1 + \beta_2 - 1 = 0$ against $H_a : \beta_1 + \beta_2 - 1 < 0$ and uses special ADF statistics, with critical values calculated by computer. If the value of ADF less than critical null hypothesis is rejected. **[5 marks]**

c) For the period II the test for OIL_BRENT Dickey-Fuller test without time trend and intercept gives significant result while the test with trend and intercept becomes insignificant with significant t-statistics for time trend. Both Dickey-Fuller test statistic and time trend becomes insignificant for the period III. What is the augmented Dickey-Fuller test with time trend and intercept? Help the student to interpret obtained results from the point of view of economics. If time trend is detected how to get rid of it?

(b) The equation with the time trend (and intercept) is

$$\Delta X_t = \beta_0 + (\beta_1 + \beta_2 - 1) X_{t-1} - \beta_2 \Delta X_{t-1} + \gamma t + u_t$$

In both cases (with and without time trend) the ADF test is based on the null *hypothesis* of no stationarity $H_0 : \beta_1 + \beta_2 - 1 = 0$ against $H_a : \beta_1 + \beta_2 - 1 < 0$ and uses special ADF statistics, with critical values calculated by computer. If the value of ADF less than critical null hypothesis is rejected.

If the null hypothesis here is rejected there is still the chance that the time series is time non-stationary, when a definite time trend violates the condition of constant expectation (one of conditions of so called weak stationarity). To test this it is sufficient to run conventional t-test for coefficient of time: $H_0 : \gamma = 0$ against $H_0 : \gamma \neq 0$.

If time trend is detected the simple procedure allows to get rid of it:

1) estimate equation $\hat{X}_t = \hat{\gamma}_0 + \hat{\gamma}_2 t$ and memorize its residuals $e_t = X_t - \hat{X}_t$;

2) use e_t instead of X_t in the model $X_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_t$ or another models under consideration.

Equivalent procedure is to include time variable explicitly into original equation

$$X_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \gamma t + u_t$$

In the period II the student observed non stationarity caused by the time trend. In the period III time trend disappears, the behavior of oil priced becomes chaotic, non stationarity now is of a difference type.

[5 marks]

d) The student also wants to investigate the causality in a pair of time series RUR_USD and OIL_BRENT using Granger causality test with two lags. He obtains the following results

Null Hypothesis:	Obs	F-Statistic	Prob.
Period I			
RUR_USD does not Granger Cause OIL_BRENT	159	1.89361	0.1540
OIL_BRENT does not Granger Cause RUR_USD		1.54372	0.2169
Period II			
RUR_USD does not Granger Cause OIL_BRENT	87	1.74592	0.1809
OIL_BRENT does not Granger Cause RUR_USD		6.95896	0.0016
Period III			
RUR_USD does not Granger Cause OIL_BRENT	26	6.74149	0.0055
OIL_BRENT does not Granger Cause RUR_USD		7.01836	0.0046

Explain what is Granger causality test in the context of the problem under consideration and help the student to interpret the results of the test.

Granger test for causality: regress current value of *X* on the past values of *X* and *Y*, and the current value of *Y* on the past values of *X* and *Y*:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \varepsilon_t \quad (1)$$

$$X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \varepsilon_t \quad (2)$$

Test statistic: F-statistic for $H_0: \beta_1 = \beta_2 = 0$. If in (1) H_0 is rejected, then *X* Granger causes *Y*.

If in (2) H_0 is rejected, then *Y* Granger causes *X*.

For the period I we can observe independence, in the period II OIL_BRENT Granger cause RUR_USD, in the period III we observe bilateral Granger causality.

Note: bonus to those students who understand that official exchange rate is set up actually not in current but in the previous period so the equations should be corrected (one need lag OIL_BRENT). The number of observation in the third period is small **[5 marks].**

9. A study of applications for home mortgages used the linear probability model

$$MORT_i = \beta_0 + \beta_1 INC_i + \beta_2 AGE_i + \beta_3 PROP_i + u_i$$

where

$MORT_i = 1$ if a mortgage is granted: 0 otherwise,

INC_i = income of applicant in £1000,

AGE_i = age of applicant in years,

$PROP_i$ = age of the property for which the mortgage is being applied,

$i = 1, 2, \ldots, 80$.

(a) The estimated coefficient for INC_i was 1.48 with standard error 0.51. Is this coefficient significant? What is the interpretation of this coefficient?

Solution and marking

a) $t = 1.48/0.51 = 2.90$ which is significantly different from 0 at any significance level as number of observation is large enough. As *INC* increases by £1000 *MORT* increases by 1.48 but since the estimated MORT can be interpreted as a probability the prediction is likely to lie outside $[0,1]$. More accurate interpretation could be achieved using some other units, for example as *INC* increases by £10 *MORT* (interpreted as probability to be granted a mortgage) increases by 0.0148, i.e. by 1.48 percentage points. **[Total 6 marks for a]**.

(b) What are the problems associated with the linear probability model?

Solution and marking

b) As (a) shows the linear probability model can produce predictions outside the range $[0,1]$ which is not feasible. Also the error terms are heteroskedastic since

$$\varepsilon_t = 1-\beta_0-\beta_1 INC-\beta_2 AGE-\beta_3 PROP \text{ with probability } p$$

and

$$\varepsilon_t = -\beta_0-\beta_1 INC-\beta_2 AGE-\beta_3 PROP \text{ with probability } (1-p).$$

The distribution of the error term is certainly not normal so usual test strictly speaking could not be applied. Additional drawback of the model is the constant marginal effect of the explanatory variable.

[Total 6 marks for b].

(c) What is a logit model? Does it overcome the problems you listed in (b)?

Solution and marking

c) The logit model refers to a model where the probability of the occurrence of the event is determined by the logistic function, $f(z)=1/(1+e^{-z})$, where $z_i=\beta'_0+\beta'_1 INC_i+\beta'_2 AGE_i+\beta'_3 PROP_i+u'_i$. The function is confined within the range $[0,1]$ as can be seen by allowing z to tend to $+\infty$ or $-\infty$ and therefore does not suffer from the same problems as the linear probability model. It can be relatively easily estimated by maximum likelihood techniques therefore the estimates have the standard maximum likelihood properties i.e. the estimators are consistent, asymptotically efficient and asymptotically normally distributed.

[Total 6 marks for c].

(d) How the impact of each factor INC_i, AGE_i and $PROP_i$ could be evaluated? What is the difference in this point as to compare to the linear probability model?

Solution and marking

d) As it has been shown in c) now $MORT$ is a function of z: $MORT(z)=1/(1+e^{-z})$ while z_i is a function of INC, AGE, and $PROP$, according to regression $z_i=\beta'_0+\beta'_1 INC_i+\beta'_2 AGE_i+\beta'_3 PROP_i+u'_i$. So the impact of each factor could be found using chain rule For example

$$\frac{\partial MORT}{\partial INC}=\frac{\partial MORT}{\partial z}\cdot\frac{\partial z}{\partial INC}=\frac{\partial 1/(1+e^{-z})}{\partial z}\cdot\beta'_1=\frac{e^{-z}}{(1+e^{-z})^2}\cdot\beta'_1.$$

As $z=\beta'_0+\beta'_1 INC+\beta'_2 AGE+\beta'_3 PROP+u'_i$ the value of this impact depends on the point (INC, AGE, $PROP$) (it is not constant now). Usually to characterize this impact average point $(\overline{INC_i}, \overline{AGE_i}, \overline{PROP_i})$ is used as the curve $f(z)=\dfrac{e^{-z}}{(1+e^{-z})^2}$ is bellshaped, but also possible to compare impact in different points.

In any case

$$ME_{INC}=\frac{e^{-\hat{\beta}'_0+\hat{\beta}'_1 INC+\hat{\beta}'_2 AGE+\hat{\beta}'_3 PROP}}{(1+e^{-\hat{\beta}'_0+\hat{\beta}'_1 INC+\hat{\beta}'_2 AGE+\hat{\beta}'_3 PROP})^2}\cdot\hat{\beta}'_1$$

$$ME_{AGE}=\frac{e^{-\hat{\beta}'_0+\hat{\beta}'_1 INC+\hat{\beta}'_2 AGE+\hat{\beta}'_3 PROP}}{(1+e^{-\hat{\beta}'_0+\hat{\beta}'_1 INC+\hat{\beta}'_2 AGE+\hat{\beta}'_3 PROP})^2}\cdot\hat{\beta}'_2$$

$$ME_{PROP}=\frac{e^{-\hat{\beta}'_0+\hat{\beta}'_1 INC+\hat{\beta}'_2 AGE+\hat{\beta}'_3 PROP}}{(1+e^{-\hat{\beta}'_0+\hat{\beta}'_1 INC+\hat{\beta}'_2 AGE+\hat{\beta}'_3 PROP})^2}\cdot\hat{\beta}'_3$$ **[Total 7 marks for d]**.

10. A student was told to take the data for 26 OECD countries for three years (2011-2013), on w_{it} (average annual percentage rate of growth of wages for country *i* during time period *t* ($t=1$ for 2011, $t=2$ for 2012, $t=3$ for 2013) and p_{it} (average annual percentage rate of growth of productivity for country *i* during time period *t*). He found all data for 2011 and 2013 but failed to find any data for 2012. Being limited if time he decided to use the mean values of 2011 and 2013 as the interpolations for 2012, namely $w_{i2}=\frac{w_{i1}+w_{i3}}{2}$, $p_{i2}=\frac{p_{i1}+p_{i3}}{2}$, $i=1,....,26$. Then he estimates Fixed Effect regression $w_{it}=\beta_1+\beta_2 p_{it}+\alpha_i+u_{it}$ $i=1,....,26;$ $t=1,2,3$. Here α_i is unobserved heterogeneity term that in Fixed Effect regression supposed to be a combined result of influence of some non-random factors.

(a) What are advantages in analysis of panel data comparing to cross-section regression and time series analysis? Explain briefly what is the Fixed Effect approach to the estimation of the Panel Model. Why here it is the only possible approach?

Solution

a) Panel regressions allows to analyze data that both vary in time (time series) and characterize many objects (cross section). The volume of data available for analysis increases enormously. It gives possibility to remove and partly identify the heterogeneity that remains unobserved in time series and cross section analysis. Unobserved heterogeneity influence disturbance term that can correlate with explanatory variable and so make estimates inconsistent and biased.

Fixed effect approach assumed that unobserved heterogeneity α_i is determined by some fixed combined result of influence of some non-random factors $\alpha_i=\sum\gamma_j Z_{ij}$. To get rid of it one can use First Differences (FD) approach

$$w_{it}-w_{it-1}=\beta_1+\beta_2(p_{it}-p_{it-1})+u_{it}-u_{it-1} \quad i=1,....,26; t=2,...$$

This approach allows to eliminate unobservable α_i, at the cost of losing 26 degrees of freedom.

Alternative Within Groups (WG) approach is based on the idea of substitution instead of observations their deviations from the group means: $w_{it}-\overline{w}_i=\beta_1+\beta_2(p_{it}-\overline{p}_i)+u_{it}-\bar{u}_i$, this approach also allows to eliminate unobservable α_i, at the cost of losing 26 degrees of freedom, both methods also eliminate constant and all factors remaining unchanged in time. The third approach for applying FE model is LSDV method where the dummies are introduced for each object. The LSDV estimates are the same as WG ones, though the standard errors are different.

Here Random Effects cannot be applied since the OECD countries cannot be considered as objects taken randomly from the same distribution (assumption 1 for RE). So the FE is the only applicable.

[5 marks] for correct explanation.

(b) The student uses for estimation panel regression $w_{it}=\beta_1+\beta_2 p_{it}+\alpha_i+u_{it}$ $i=1,....,26;$ $t=1,2,3$ Within Group (WG) and First Difference (FD) methods. Explain formally why the coefficient estimates in the Within Groups and First Differences regressions are the same.

Solution

b) The original data for the situation under consideration can be represented by two tables (26 observations of E and corresponding 26 observations of G for two years)

$$\begin{pmatrix} w_{1,1} & \cdots & w_{25,1} \\ \frac{w_{1,1}+w_{1,3}}{2} & \cdots & \frac{w_{25,1}+w_{25,3}}{2} \\ w_{1,3} & \cdots & w_{25,3} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} p_{1,1} & \cdots & p_{25,1} \\ \frac{p_{1,1}+p_{1,3}}{2} & \cdots & \frac{p_{25,1}+p_{25,3}}{2} \\ p_{1,3} & \cdots & p_{25,3} \end{pmatrix}$$

When first difference method is applied each two observations are replaced by their first differences so the data now is

$$\begin{pmatrix} w_{1,1}-\dfrac{w_{1,1}+w_{1,3}}{2} & \cdots & w_{26,1}-\dfrac{w_{26,1}+w_{26,3}}{2} \\ \dfrac{w_{1,1}+w_{1,3}}{2}-w_{1,3} & \cdots & \dfrac{w_{26,1}+w_{26,3}}{2}-w_{26,3} \end{pmatrix} \text{ and } \begin{pmatrix} p_{1,1}-\dfrac{p_{1,1}+p_{1,3}}{2} & \cdots & p_{26,1}-\dfrac{p_{26,1}+p_{26,3}}{2} \\ \dfrac{p_{1,1}+p_{1,3}}{2}-p_{1,3} & \cdots & \dfrac{p_{26,1}+p_{26,3}}{2}-p_{26,3} \end{pmatrix}. \text{ But}$$

this is just the same as $\begin{pmatrix} \dfrac{w_{1,1}-w_{1,3}}{2} & \cdots & \dfrac{w_{26,1}-w_{26,3}}{2} \\ \dfrac{w_{1,3}-w_{1,1}}{2} & \cdots & \dfrac{w_{26,3}-w_{26,1}}{2} \end{pmatrix}$ and $\begin{pmatrix} \dfrac{p_{1,1}-p_{1,3}}{2} & \cdots & \dfrac{p_{26,1}-p_{26,3}}{2} \\ \dfrac{p_{1,3}-p_{1,1}}{2} & \cdots & \dfrac{p_{26,3}-p_{26,1}}{2} \end{pmatrix}$ (*)

When within group method is applied each three observations are transformed into the difference between each of them and their mean, for example $\begin{matrix} w_{i,1} \\ \dfrac{w_{i,1}+w_{i,3}}{2} \\ w_{i,3} \end{matrix} \rightarrow \begin{matrix} w_{i,1}-\dfrac{w_{i,1}+w_{i,3}}{2} \\ \dfrac{w_{i,1}+w_{i,3}}{2}-\dfrac{w_{i,1}+w_{i,3}}{2} \\ w_{i,3}-\dfrac{w_{i,1}+w_{i,3}}{2} \end{matrix}$.

But this is simply $\begin{matrix} \dfrac{w_{i,1}-w_{i,3}}{2} \\ 0 \\ \dfrac{w_{i,3}-w_{i,1}}{2} \end{matrix}$.

The same with p, so the data now is

$$\begin{pmatrix} \dfrac{w_{1,1}-w_{1,3}}{2} & \cdots & \dfrac{w_{26,1}-w_{26,3}}{2} \\ 0 & \cdots & 0 \\ \dfrac{w_{1,3}-w_{1,1}}{2} & \cdots & \dfrac{w_{26,3}-w_{26,1}}{2} \end{pmatrix} \text{ and } \begin{pmatrix} \dfrac{p_{1,1}-p_{1,3}}{2} & \cdots & \dfrac{p_{26,1}-p_{26,3}}{2} \\ 0 & \cdots & 0 \\ \dfrac{p_{1,3}-p_{1,1}}{2} & \cdots & \dfrac{p_{26,3}-p_{26,1}}{2} \end{pmatrix} \text{ (**)}$$

Comparing (*) and (**) we can see that all non-zero observations are the same. So the slope coefficient which is here $b=\dfrac{\sum w_i p_i}{\sum p_j^2}$ (the formula for the model without intercept) will not change.

<u>Marking</u>

[8 marks] for correct explanation.

(c) What are comparative advantages and disadvantages of Within group and First difference methods in estimating panel regressions? Why the application of any of these methods leads to autocorrelation problem? How serious is this problem in each of these methods? How it could be mitigated?

<u>Solution</u>

c) The main advantage: the FE methods can describe the unobserved heterogeneity, and hence to get unbiased and consistent estimates.

Disadvantages:

- Loss of n degrees of freedom;
- Loss of the intercept;
- Loss of all the variables which are constant in time for each unit;
- Less precise estimates due to smaller variation of explanatory variables;
- Autocorrelation in the disturbance term leading to inefficiency.

When applying First Differences (FD) approach

$$w_{it}-w_{it-1}=\beta_1+\beta_2(p_{it}-p_{it-1})+u_{it}-u_{it-1} \quad i=1,\ldots,26;\ t=1,\ldots,T.$$

the disturbance term reveals autocorrelation of the moving average type

$$\varepsilon_{it} = u_{it} - u_{it-1}$$

The lagged disturbance term is

$$\varepsilon_{it-1} = u_{it-1} - u_{it-2}$$

so the successive values of the disturbance term for the same unit become correlated

$$\mathrm{cov}(\varepsilon_{it};\varepsilon_{it-1}) = \mathrm{cov}(u_{it} - u_{it-1}; u_{it-1} - u_{it-2}) = \mathrm{cov}(u_{it}; u_{it-1}) - \mathrm{cov}(u_{it-1}; u_{it-1}) - \mathrm{cov}(u_{it}; u_{it-2}) + \mathrm{cov}(u_{it-1}; u_{it-2}) =$$

$$= 0 - \sigma_u^2 - 0 \ + 0 \quad = -\sigma_u^2$$

So $\mathrm{cov}(\varepsilon_{it};\varepsilon_{it-1}) = -\sigma_u^2$ (*)

This approach allows to eliminate unobservable α_i, at the cost of losing 26 degrees of freedom.

Alternative Within Groups (WG) approach is based on the idea of substitution instead of observations their deviations from the group means:

$$w_{it} - \overline{w_i} = \beta_1 + \beta_2(p_{it} - \overline{p}_i) + u_{it} - \overline{u}_i$$

The disturbance term here is autocorrelated for all observations belonging to the same unit:

$$\varepsilon_{it1} = u_{it1} - \overline{u_i}$$

and

$$\varepsilon_{it2} = u_{it2} - \overline{u_i}$$

They are correlated: $\mathrm{cov}(\varepsilon_{it1};\varepsilon_{it2}) = \mathrm{cov}(u_{it1} - \overline{u}_i; u_{it2} - \overline{u}_i) = \mathrm{cov}(u_{it1} - \frac{\sum_t u_{it}}{T}; u_{it2} - \frac{\sum_t u_{it}}{T})$,

dropping zero terms ($\mathrm{cov}(u_{it};u_{is}) = 0, t \neq s$) we get

$$\mathrm{cov}(\varepsilon_{it1};\varepsilon_{it2}) = -\mathrm{cov}(u_{it1}; \frac{u_{it1}}{T}) - \mathrm{cov}(u_{it2}; \frac{u_{it2}}{T}) + \mathrm{cov}(\frac{u_{it}}{T}; \frac{u_{it}}{T}) + \sum_{\tau} \ \mathrm{cov}(\frac{u_{i\tau}}{T}; \frac{u_{i\tau}}{T}) = \frac{-2\sigma_u^2}{T} + \frac{\sigma_u^2}{T^2}T$$

So $\mathrm{cov}(\varepsilon_{it1};\varepsilon_{it2}) = \frac{-\sigma_u^2}{T}$ (**)

Comparing with (*) we can see that in WG method the problem of induced autocorellation is not so severe as in the FD, but still persists.

For the mitigation of the autocorrelation problem Autoregressive transformation (special version) can be applied. Also the LSDV method can be used which does not have autocorrelation in the disturbance term.

Marking

[7 marks] for correct explanation.

The International College of Economics and Finance
Econometrics – 2015-2016.
Midterm exam 2016. March 31.

SOLUTION&MARKING

1. Consider a simple linear regression model:

$$Y_i = \beta_0 + \beta_1 X_i + u_i,\ i = 1, 2, ..., n \qquad (1)$$

where $\mathrm{E}(u_i) = 0$; $\mathrm{E}(u_i^2) = \sigma^2$ and $\mathrm{E}(u_i u_j) = 0$ if $i \neq j$, X_i assumed to be non-stochastic.

(a) Suppose that the fitted line is $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_i$ where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the OLS estimators. Prove that the fitted line must pass through the point $(\overline{X}, \overline{Y})$ representing the mean of the variables in the sample. The student believes that the theoretical regression (1) also must satisfy this condition. Is he right? Explain.

Solution: Since $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$ (one of two least square equations), rearranging

$$\overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{X}$$

and so $(\overline{Y}, \overline{X})$ lies on the regression line.

We cannot say the same about theoretical regression, as the presence of unobservable disturbance terms makes this equality possible only when by accidence $\overline{u} = 0$

$$\overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{X} + \overline{u}$$

[5 marks]

МАРКИНГ: Все говорили, что этот вопрос слишком простой, но на деле с ним мало кто хорошо справился.

1) Нужно четко понимать, что равенство $\overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{X}$ не нуждается в доказательстве, поскольку представляет собой просто одно из уравнений метода наименьших квадратов. Разумеется, студенты могут его вывести из принципа наименьших квадратов, но больше баллов за это они не получат, так как хорошо известно, что начальство нужно знать в лицо. Так что за первую часть я бы смело ставил 3 балла.

2) Второе утверждение тоже одно из базовых, путаница может происходить из-за того, что сумма остатков по выборке всегда равна нулю. Поэтому нужно проверить, что студент четко понимает различие между остатками и случайным членом. Разумеется, в явной форме про остатки здесь он говорить не обязан. 2 балла.

(b) An investigator correctly believes that the relationship between the variables *X* and *Y* is described by the linear model specified above. Given a sample of *n* observations, the investigator estimate β_1 by calculating it as the average value of *Y* divided by the average value of *X*: $\hat{\beta}_1 = \dfrac{\overline{Y}}{\overline{X}}$.

Discuss the properties of this estimator. What difference would it make if it could be assumed that $\beta_0 = 0$?

Solution: Since

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$
$$\overline{Y} = \beta_0 + \beta_1 \overline{X} + \overline{u}$$

and

$$\hat{\beta}_1 = \frac{\overline{Y}}{\overline{X}} = \frac{\beta_0 + \beta_1 \overline{X} + \overline{u}}{\overline{X}} = \frac{\beta_0}{\overline{X}} + \beta_1 + \frac{\overline{u}}{\overline{X}}.$$

Hence, assuming that *X* is non-stochastic,

$$\mathrm{E}(\hat{\beta}_1) == \frac{\beta_0}{\overline{X}} + \beta_1 + \frac{1}{\overline{X}}\mathrm{E}(\overline{u}) = \frac{\beta_0}{\overline{X}} + \beta_1$$

since $\mathrm{E}(\bar{u})=0$. Thus $\hat{\beta}_1$ is biased unless $\beta_0=0$, and the direction of the bias depends on the sign of both β_0 and $\bar{X}$. Since

$$\lim_{n\to\infty}\mathrm{E}(\hat{\beta}_1)\neq\beta_1,$$

the estimator is not consistent.

Special case: $\beta_0=0$. $\hat{\beta}_1$ is now unbiased and $\mathrm{var}(\hat{\beta}_1)=\mathrm{var}(\frac{\bar{Y}}{\bar{X}})=\mathrm{var}(\frac{\frac{1}{n}\sum_{j=1}^{n}Y_i}{\bar{X}})=\frac{1}{n^2}\frac{n\sigma^2}{(\bar{X})^2}=\frac{\sigma^2}{n(\bar{X})^2}\to 0$ as $n\to\infty$. Hence if $\beta_0=0$, then $\hat{\beta}_1$ is a consistent estimator of β_1.

In comparison to the OLS estimator $\hat{\beta}_1^{OLS}=\frac{\sum X_iY_i}{\sum X_i^2}$, $\hat{\beta}_1$ is inefficient. If the Gauss-Markov assumptions hold, the OLS estimator is the most efficient estimator.

[5 marks]

МАРКИНГ: Дима Малахов совершенно справедливо говорит, что понятие состоятельности применительно к конечной нестохастической выборке неприменимо, я тоже так учу студентов, но англичане с нами не согласны (и я всегда это говорю студентам), и особенно не запариваются насчет точного смысла предпосылок разных моделей A, B, C. В общем, они правы, все эти теоретические схемы зависят от точки зрения, а состоятельность выражает простую мысль, что если «долго мучиться – что-нибудь получится», то есть увеличение выборки дает выигрыш и в принципе приводит к желаемому результату точной оценки. Поэтому я не стали ничего менять в лондонской маркинг-схеме.

1) В проверке свойств важен порядок: сначала смещение, потом несостоятельность, потом, если уже все это наладили, начинаем обсуждать эффективность.

2) У предложенной оценки ясный геометрический смысл – наклон прямой, проведенной из начала координат в точку средних. Поэтому обсуждение очень удобно сопровождать графиками. Тем, кто пользуется графиками, можно добавить балл в пределах общей пятерки.

(c) Is the OLS estimator of β_1 consistent? Explain in detail.
Is it efficient? What difference would it make if it could be assumed that $\beta_0=0$?

Solution and marking

c) OLS estimator of OLS is $\hat{\beta}_1^{OLS}=\frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{X})_i^2}=\beta_1+\frac{\sum(X_i-\bar{X})u_i}{\sum(X_i-\bar{X})_i^2}$, so

$E(\hat{\beta}_1^{OLS})=\beta_1+\frac{\sum(X_i-\bar{X})Eu_i}{\sum(X_i-\bar{X})_i^2}=\beta_1$ and $\mathrm{var}(\hat{\beta}_1^{OLS})=\frac{\sum(X_i-\bar{X})^2\,\mathrm{var}(u_i)}{\left(\sum(X_i-\bar{X})_i^2\right)^2}=\frac{\sigma^2}{\sum(X_i-\bar{X})^2}$.

It follows that $\lim_{T\to\infty}V(\hat{\beta}_1^{OLS})\to 0$ since as the sample size increases $\sum(X_i-\bar{X})^2$ must increase.

Hence a sufficient condition for consistency holds $\Rightarrow$ $\hat{\beta}_1^{OLS}$ is a consistent estimator of β_1.

Gauss-Markov theorem states that OLS estimator $\hat{\beta}_1^{OLS}$ is the efficient estimator. The situation changes if it could be assumed that $\beta_0=0$: now OLS estimator is $\hat{\beta}_1^{OLS}=\frac{\sum X_iY_i}{\sum X_{ii}^2}$ so the estimator $\frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{X})_i^2}$ becomes inefficient.

[5 marks]

МАРКИНГ: Я пока написал здесь максимум возможного, чтобы легче было проверять. Даже чуть ниже привел разные альтернативные способы вывода. Потом в решении для студентов сократим. Самое главное проверить, что студенты поняли все части вопроса, на каждую из них ответили, и что они понимают о какой оценке они говорят, а не просто произносят заклинания про OLS и Гаусса-Маркова

ВОТ ЕЩЕ ПРЯМОЕ ДОКАЗАТЕЛЬСТВО ИЗ ДРУГОЙ ЛОНДОНСКОЙ МАРКИНГ-СХЕМЫ

Explain what is meant by consistency in a statistical estimator. Under what conditions the least squares estimate of the slope coefficient in a simple regression of y_t on x_t is consistent?

Solution and marking

The definition and sufficient condition of consistency should be given. It should be illustrated with an example. The solution follows.

If $\hat{\theta}$, based on a sample of size *T*, is a consistent estimator of θ then $\Pr(|\hat{\theta}-\theta|>e)\to 0$ as $T\to\infty$ for every $e>0$. Another way of expressing this is that $\hat{\theta}$ converges in probability to θ. In short, we can write the above statement as $\text{plim}\,\hat{\theta}=\theta$, where plim stands for the probability limit. Hence if $\text{plim}\,\hat{\theta}=\theta$ then $\hat{\theta}$ is a consistent estimator of θ.

The sufficient condition for consistency is

$$E(\hat{\theta})=\theta$$

or

$$\lim_{T\to\infty} E(\hat{\theta})=\theta \text{ and } \lim_{T\to\infty}\text{Var}(\hat{\theta})=\theta .$$

Let the model be

$$y_t=\beta x_t+u_t,\ t=1,2,\dots,T$$

$E(u_t)=0$, $E(u_t^2)=\sigma^2$ and $E(u_s u_t)=0$ if $s\neq t$, for all $s,t=1,2,\dots,T$.

$$\hat{\beta}=\frac{\sum x_t y_t}{\sum x_t^2}=\beta+\frac{\sum x_t u_t}{\sum x_t^2}$$

$$\text{plim}(\hat{\beta})=\beta+\frac{\text{plim}(\sum x_t u_t)/T}{\text{plim}(\sum x_t^2)/T}=\beta+\frac{\text{cov}(x,u)}{\text{var}(x)}=\beta+\frac{0}{\sigma_x^2}=\beta .$$

Hence $\hat{\beta}$ is a consistent estimator of β.

Assumptions: The *x*s are non-stochastic, or $\text{cov}(x,u)=0$.

А ВОТ АЛЬТЕРНАТИВНЫЙ ВЫВОД ДЛЯ ДИСПЕРСИИ

$$\text{var}(\hat{\beta})=\text{E}(\hat{\beta}-\beta)^2=\text{E}\left(\frac{\sum_{t=1}^{n}x_t u_t}{\sum_{t=1}^{n}x_t^2}\right)^2=\frac{\text{E}(x_1^2u_1^2+x_2^2u_2^2+x_3^2u_3^2+\dots+2x_1x_2u_1u_2+\dots)}{\left(\sum_{t=1}^{n}x_t^2\right)^2}$$

Now we have assumed that $\text{E}(u_t^2)=\sigma^2$ and $\text{E}(u_t u_s)=0$ for $t\neq s$ so

$$\text{var}(\hat{\beta})=\frac{\sigma^2\sum_{t=1}^{n}x_t^2}{\left(\sum_{t=1}^{n}x_t^2\right)^2}=\frac{\sigma^2}{\sum_{t=1}^{n}x_t^2}.$$

2. A researcher decided to investigate the relationship between the economic growth and the development of financial sector in the country R. The main variables of interest were X_t - the per capita real GDP and Y_t - the ratio of the credit to private sector to the GDP. The data for 45 years was available. Both variables were taken in the logarithmic form (as $\log Y_t$ and $\log X_t$). The variable Y_t was used as the proxy for the development of financial sector in the country. The variables I_t (investment-to-GDP ratio) and H_t (the combined indicator of human capital) were used as the control variables.

a) Explain how the variable $\log X_t$ may be used for measuring economic growth. Explain the role of proxy variables in economic analysis. Explain the role of the control variables in the model and potential consequences of their omission.
The researcher started with the investigation of Granger causality between $\log Y_t$ and $\log X_t$. He found that in the country R $\log X_t$ Granger causes $\log Y_t$, while $\log Y_t$ does not Granger cause $\log X_t$, with the maximum lag equal 2. Describe the model used for Granger causality investigation, the tests done and their results which led to the researcher's conclusion.

Solution

a) The growth rates can be measured as a difference of $\log X_t$: $\log\left(\frac{X_t}{X_{t-1}}\right) = \log X_t - \log X_{t-1}$.

Proxy variables are used when the data on some regressor are not available. If you just ignore this variable, the regression results of the estimators of coefficients of all other variables in equation will be biased (omitted variable bias). If proxy variable correlates with omitted one (and not correlates with disturbance term) the bias disappears or becomes smaller. However the coefficient of the proxy variable cannot be used as substitute for the coefficient of the omitted variable. The variables included into equation (regressors) are called control variables, as the main reason of their presence in equation is not their significance or meaning but to prevent omitted variable bias that would occur in their absence.

If in the country R $\log X_t$ Granger causes $\log Y_t$ (maximum lag = 2) means that in equation $\log Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \beta_1 X_{t-1} + \beta_2 X_{t-2} + u_t$ the hypothesis $H_0: \beta_1 = \beta_2 = 0$ is rejected using appropriate F-statistic. Conversely $\log Y_t$ does not Granger causes $\log X_t$ means that in equation $\log X_t = \gamma_0 + \gamma_1 X_{t-1} + \gamma_2 X_{t-2} + \delta_1 Y_{t-1} + \delta_2 Y_{t-2} + v_t$ the hypothesis $H_0: \delta_1 = \delta_2 = 0$ is not rejected using appropriate F-statistic. It has nothing to do with cause and effect relationship, but states informational input of the variable $\log X_t$ in explanation of the variance of the dependent variable $\log Y_t$. Here you are an example of unidirectional Granger causality.
[5 marks]
МАРКИНГ: Проверить что студент ответил на все вопросы. Я специально четко структурировал ответы по трем частям. Кажется, делить нужно в пропорции 1+1+3 или 0,5+1,5+3.

b) Then the researcher tested the variables $\log Y_t$, $\log X_t$, I_t and H_t for stationarity using the ADF test with lag=1. He found that the variables $\log Y_t, \log X_t$ and H_t are I(1), while I_t is I(0). Having that, the researcher regressed $\log Y_t$ on $\log X_t$ and H_t and found that they are cointegrated. Then he calculated the first differences $\Delta \log Y_t$, $\Delta \log X_t$ and ΔH_t, and fitted the regression of $\Delta \log Y_t$ on $\Delta \log X_t$, ΔH_t, I_t and the lagged residuals of cointegrating relationship. Describe the typical model used for the test, and the procedure of test implementation.

Solution
a) Here we have two tests: Dickey-Fuller test for stationary, and test for cointegration. To test, for example, variable $\log Y_t$ for stationarity one should run equation (standard Dickey and Fuller test for a unit root) and is based on the model $\log Y_t = \beta_1 + \beta_2 \log Y_{t-1} + \gamma t + u_t$ which can be re-written as

$\Delta \log Y_t = \beta_1 + (1-\beta_2)\log Y_{t-1} + \gamma t + u_t$ where $\Delta \log Y_t = \log Y_t - \log Y_{t-1}$. The null hypothesis for stationarity is $H_0: 1-\beta_2 = 0$, $H_A: 1-\beta_2 \neq 0$. We cannot use the standard *t*-test procedure in this case because the distribution of the *t*-statistic is not a *t*-distribution so critical values have been computed by Dickey and Fuller using Monte-Carlo techniques. Additionally time series can be tested for the presence of the time trend – we use conventional *t*-test for the coefficient γ of the variable t.

Dickey Fuller test is sensitive to the presence of serial correlation in the error term so we need to take steps to remove the effects of this serial correlation – this is done by including additional lagged values of y_t in the regression, i.e. $\log Y_t = \beta_1 + \beta_2 \log Y_{t-1} + \beta_3 \log Y_{t-2} + \gamma t + u_t$ for an $AR(1)$ serial correlation. Then after some transformations

$\log Y_t - [\log Y_{t-1}] = \beta_1 + \beta_2 \log Y_{t-1} - [\log Y_{t-1}] + (\beta_3 \log Y_{t-1} - \beta_3 \log Y_{t-1}) + \beta_3 \log Y_{t-2} + \gamma t + u_t$

$\Delta \log Y_t = \beta_1 + \beta_2 \log Y_{t-1} + \beta_3 \log Y_{t-1} - \log Y_{t-1} - (\beta_3 \log Y_{t-1} + \beta_3 \log Y_{t-2}) + \gamma t + u_t$ we get

$\Delta \log Y_t = \beta_1 + (\beta_2 + \beta_3 - 1)\log Y_{t-1} - \beta_3 \Delta \log Y_{t-1} + \gamma t + u_t$ with null hypothesis $H_0: 1 - \beta_1 - \beta_2 = 0$ (augmented Dickey-Fuller test). Once again using Dickey-Fuller tables.

To test a group of time series for cointegration we have first to run cointegration relationship $\log Y_t = \beta_1 + \beta_2 \log X_t + \beta_3 H_t + u_t$ (it is said that it exists), and finding estimates of coefficients evaluate the residuals: $e_t = \log Y_t - \hat{\beta}_1 - \hat{\beta}_2 \log X_t - \hat{\beta}_3 H_t$. Now these residuals should be tested for stationarity using Dickey Fuller test (without constant and time trend, based on the properties of residuals). Now first difference model can be enriched by the obtained residuals ($\Delta \log Y_t$ on $\Delta \log X_t$, ΔH_t, I_t and the lagged residuals e_{t-1} are all stationary) to get a model, combined short-term dynamics of the differences with the long-term dynamics of integrating relationship.

[5 marks]

МАРКИНГ: Ответ разбивается на две четких части, первая просто тест Дики-Фуллера (наверное, 1 + 2 за модель с дополнительным лагом), и потом 2 за тестирование коинтеграции.

c) Demonstrate the ADL(1,1) model which led to the last regression. Explain how you estimate the short-run and the long-run elasticities of Y_t by X_t. Explain how the error correction mechanism works.

Solution

a) Consider ADL(1,1) model of the type $\log Y_t = \beta_1 + \beta_2 \log Y_{t-1} + \beta_3 \log X_t + \beta_4 \log X_{t-1} + u_t$ (for simplicity, we restrict ourselves to the two variables). The coefficient β_3 is the short-run elasticity of Y_t by X_t (if Y_{t-1} and X_{t-1} are fixed, $d \log Y_t = \frac{dY_t}{Y_t} = \beta_3 \frac{dX_t}{X_t}$, so $\beta_3 = \frac{\frac{dY_t}{Y_t} \cdot 100\%}{\frac{dX_t}{X_t} \cdot 100\%}$).

In equilibrium we would have the relationship $\overline{\log Y} = \beta_1 + \beta_2 \overline{\log Y} + \beta_3 \overline{\log X} + \beta_4 \overline{\log X}$, so $(1-\beta_2)\overline{\log Y} = \beta_1 + (\beta_3 + \beta_4)\overline{\log X}$

and $\overline{\log Y} = \frac{\beta_1}{1-\beta_2} + \frac{\beta_3 + \beta_4}{1-\beta_2}\overline{\log X}$. Thus $\frac{\beta_3 + \beta_4}{1-\beta_2}$ is a long-run elasticity of Y_t by X_t.

This is the cointegrating relationship. The ADL(1,1) relationship may be rewritten to incorporate this relationship. First we subtract $\log Y_{t-1}$ from both sides

$\log Y_t - \log Y_{t-1} = \beta_1 + \beta_2 \log Y_{t-1} - \log Y_{t-1} + \beta_3 \log X_t + \beta_4 \log X_{t-1} + u_t$

$\Delta \log Y_t = \beta_1 + (\beta_2 - 1)\log Y_{t-1} + \beta_3 \log X_t + \beta_4 \log X_{t-1} + u_t$

Then we add $\beta_3 \log X_{t-1}$ to the right side and subtract it again (and rearrange):

$\Delta \log Y_t = \beta_1 + (\beta_2 - 1)\log Y_{t-1} + \beta_3 \log X_{t-1} + \beta_4 \log X_{t-1} + \beta_3 \log X_t - \beta_3 \log X_{t-1} + u_t$

and finally

$$\Delta\log Y_t = (\beta_2 - 1)\left(\log Y_{t-1} - \frac{\beta_1}{1-\beta_2} - \frac{\beta_3+\beta_4}{1-\beta_2}\log X_{t-1}\right) + \beta_3(\log X_t - \log X_{t-1}) + u_t$$

Hence we obtain a model that states that the change in $\log Y_t$ in any period will be governed by the change in $\log X_t$ and the discrepancy between $\log Y_{t-1}$ and the value predicted by the cointegrating relationship.

The first term is denoted the error-correction mechanism. The effect of this term is to reduce the discrepancy between $\log Y_{t-1}$ and its cointegrating level. The size of the adjustment is proportional to the discrepancy.

Although the variables $\log Y_t$ and $\log X_t$ are both I(1), all of the terms in the regression equation are I(0) and hence the model may be fitted using least squares in the standard way.

Of course, the β parameters are not known and the cointegrating term is unobservable.

One way of overcoming this problem, known as the Engle–Granger two-step procedure, is to use the values of the parameters estimated in the cointegrating regression to compute the cointegrating term.

It can be demonstrated that the estimators of the coefficients of the fitted equation will have the same properties asymptotically as if the true values had been used.

[5 marks]

МАРКИНГ: Опять написано максимальное количество текста, чтобы было легче сравнивать с эталоном. Для студентов потом упростим. Самое важное – смысл разных эластичностей (вывод не обязателен), и потом вывод самой модели с коррекцией ошибок.

3. Let a linear model $Y_t = \beta_1 + \beta_2 X_t + u_t$ with the autocorrelated errors $u_t = \rho u_{t-1} + v_t$ is considered, where v has zero mean, constant variance and zero autocovariance.

a) What is meant by a common factor test in the context of a given model? How would you perform a common factor test and what hypothesis would you be testing?

Solution: It has to be shown that the $AR(1)$ model is a restricted version of the $ADL(1,1)$ model. Restrictions should be explicitly derived and the test statistic defined. The answer is:
Given the equations $Y_t = \beta_1 + \beta_2 X_t + u_t$ and $u_t = \rho u_{t-1} + v_t$, where v has zero mean, constant variance and zero autocovariance, combine the two equations (1) and (2)

$$Y_t = \beta_1 + \beta_2 X_t + \rho u_{t-1} + v_t \qquad (1)$$

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \quad | \cdot \rho$$

$$\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho u_{t-1} \qquad (2)$$

to give

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2 X_t - \rho\beta_2 X_{t-1} + v_t \text{ and}$$

$$Y_t = \beta_1(1-\rho) + \rho Y_{t-1} + \beta_2 X_t - \rho\beta_2 X_{t-1} + v_t \text{ or simply}$$

$$Y_t = \lambda_1 + \lambda_2 Y_{t-1} + \lambda_3 X_t - \lambda_2\lambda_3 X_{t-1} + v_t,$$

which is the restricted version of the general form (an ($ADL(1,1)$) model)

$$Y_t = \lambda_1 + \lambda_2 Y_{t-1} + \lambda_3 X_t + \lambda_4 X_{t-1} + v_t$$

and is subject to the restriction $\lambda_4 = -\lambda_2\lambda_3$. The test of this restriction is the common factor test.

Я ПЕРЕДЕЛАЛ РЕШЕНИЕ И ОБНАРУЖИЛ ОШИБКИ В ПОСЛЕДНИХ ДВУХ СТРОЧКАХ ЛОНДОНСКОЙ СХЕМЫ В ЗАПИСИ ОГРАНИЧЕННОЙ МОДЕЛИ, ТАК ЧТО И ЕЕ НУЖНО ВЫВЕРИТЬ!!!

Note that the usual F-test of the restriction is not appropriate because the restriction is non-linear so we have to use the test statistic

$$n\log\left(\frac{RSS_R}{RSS_U}\right) \sim \chi_1^2,$$

where RSS_R and RSS_U are the residual sum of squares from the restricted and unrestricted models respectively, n is the sample size and the test statistic is asymptotically chi-square with one degree of freedom.
[5 marks]

МАРКИНГ: Тут простая и жесткая схема – пляшем от ADL. Слова тут ничего не дают.

b) Consider now a linear model with two explanatory variables $Y_t = \beta_1 + \beta_2 X_t + \beta_3 Z_t + u_t$ with the autocorrelated errors $u_t = \rho u_{t-1} + v_t$. Derive specification of the ADL model and explain what changes in common factor test and corresponding hypotheses?

Solution: We start again from $Y_t = \beta_1 + \beta_2 X_t + \beta_3 Z_t + \rho u_{t-1} + v_t$ and subtracting lagged model multiplied by ρ:

$$\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho\beta_3 Z_{t-1} + \rho u_{t-1} \text{ we get}$$

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2 X_t - \rho\beta_2 X_{t-1} + \beta_3 Z_t - \rho\beta_3 Z_{t-1} + v_t \text{ or}$$

$$Y_t = \beta_1(1-\rho) + \rho Y_{t-1} + \beta_2 X_t - \rho\beta_2 X_{t-1} + \beta_3 Z_t - \rho\beta_3 Z_{t-1} + v_t \text{ or simply}$$

$$Y_t = \lambda_1 + \lambda_2 Y_{t-1} + \lambda_3 X_t - \lambda_2\lambda_3 X_{t-1} + \lambda_4 Z_t - \lambda_2\lambda_4 Z_{t-1} + v_t,$$

which is also the restricted version of the general form (an ($ADL(1,1)$) model)

$$Y_t = \lambda_1 + \lambda_2 Y_{t-1} + \lambda_3 X_t + \lambda_4 X_{t-1} + \lambda_5 Z_t + \lambda_6 Z_{t-1} + v_t$$

with two restrictions $\lambda_4 = -\lambda_2\lambda_3$, $\lambda_6 = -\lambda_2\lambda_4$. The test of this restriction is the common factor test $n\log\left(\frac{RSS_R}{RSS_U}\right)$ but now it follows chi-square distribution $\sim \chi_2^2$ with two degrees of freedom. Here we can clearly see the meaning of the name 'common factor': two restrictions $\lambda_4 = -\lambda_2\lambda_3$, $\lambda_6 = -\lambda_2\lambda_4$ contain common factor λ_2 which is equal to ρ in original specification. **[5 marks]**

МАРКИНГ: собственно, здесь и становится ясно, при чем тут общий множитель. Ограничения должны быть четко выписаны, тестовая статистика тоже – проверить числитель и знаменатель, число степеней свободы по числу ограничений.

c) Sometimes a common factor test is considered as the test for autocorrelation, and corresponding ADL model is considered as a tool allowing to eliminate or at least mitigate the consequences of autocorrelation. Comment.

Solution: Specification of ADL(1,1) model $Y_t = \beta_1(1-\rho) + \rho Y_{t-1} + \beta_2 X_t - \rho\beta_2 X_{t-1} + v_t$ is based on the assumption of autocorrelated disturbance term $u_t = \rho u_{t-1} + v_t$, and valid restriction $\lambda_4 = -\lambda_2\lambda_3$ means that there is no difference between restricted and unrestricted model, so if the null hypothesis is not rejected by common factor test the restricted version should be chosen as more efficient for estimation. This restricted version $Y_t = \beta_1(1-\rho) + \rho Y_{t-1} + \beta_2 X_t - \rho\beta_2 X_{t-1} + v_t$ could be decomposed again into a linear model $Y_t = \beta_1 + \beta_2 X_t + u_t$ with the autocorrelated errors $u_t = \rho u_{t-1} + v_t$, so the common factor test could reveal the autocorrelation of the first order. The coefficient ρ of Y_{t-1} is interpreted in ADL model as autocorrelation coefficient. On the other hand if common factor test allows to reject restriction(s), this coefficient is interpreted in different way simply as influence of 'habits' (for example in case of consumption of a certain good as dependent variable it is propensity to reproduce past consumer behavior).

The disturbance term v_t of both restricted and unrestricted models

$$Y_t = \lambda_1 + \lambda_2 Y_{t-1} + \lambda_3 X_t - \lambda_2\lambda_3 X_{t-1} + v_t, \text{ and}$$

$$Y_t = \lambda_1 + \lambda_2 Y_{t-1} + \lambda_3 X_t + \lambda_4 X_{t-1} + v_t$$

is free from autocorrelation, so the correct assumption on the exact value of ρ allows to eliminate autocorrelation while the knowledge of its appoximation allows only to mitigate the problem of autocorrelation (and its consequences). **[5 marks]**

МАРКИНГ: Тут нет жесткой схемы, в общем-то любые указания на то, что раз в предпосылках была автокорреляция, то ее мы и проверяем, годятся. Так что придется проследить логику отвечающего – она должна быть. Механически заученные слова тут не годятся. Я бы за все эти пояснения ставил 3 балла, а еще добавлял бы за различие в интерпретации коэффициента в модели ADL.

МАРКИНГ: Решение очень длинное, для публикации его нужно будет сократить, оставив самое главное. Но для проверки подробное решение лучше, мало ли чего напишут студенты. Вон, даже стихи написал Арсен.

4. A researcher has data on *G*, the average annual rate of growth of GDP 2008–2015, and *S*, the average years of schooling of the workforce in 2015, for 28 European Union countries. She believes that *G* depends on *S* and on *E*, the level of entrepreneurship in the country, and a disturbance term *u*:

$$G = \beta_1 + \beta_2 S + \beta_3 E + u \qquad (1)$$

u may be assumed to satisfy the usual regression model assumptions. For this purpose *S* and *E* may be treated as nonstochastic variables.
Unfortunately the researcher does not have data on *E*. The researcher does some more research and obtains data on G^*, the average annual rate of growth of GDP 2001–2007, and S^*, the average years of schooling of the workforce in 2007, for the same countries. She thinks that she can deal with the unobservable variable problem by regressing ΔG, the change in *G*, on ΔS, the change in *S*, where

$$\Delta G = G - G^*$$
$$\Delta S = S - S^*$$

assuming that *E* would be much the same for each country in the two periods. She fits the equation

$$\Delta G = \delta_1 + \delta_2 \Delta S + w \qquad (2)$$

where *w* is a disturbance term that satisfies the usual regression model assumptions.

(a) Compare the properties of the estimators of the coefficient of *S* in (1) and of the coefficient of ΔS in (2). Which model is preferable considering the pros and cons?

Solution and marking

a) The data considered here constitute a panel, as it comprises data on 28 countries for at least two time periods. The problem considered here is typical for panel data model: $G_{it} = \beta_1 + \beta_2 S_{it} + \alpha_i + u_{it}$ unobserved heterogeneity α_i is present in the model associated with some omitted variables, for example *E*, the level of entrepreneurship is mentioned $G = \beta_1 + \beta_2 S + \beta_3 E + u$. There are some methods to remove unobserved heterogeneity, one of them is differencing - first difference (FD) method, - that allows to remove from the model such unobserved factors as well as constant and all elements not depending on time.
Given (1), the differenced model should have been

$$\Delta G = \delta_2 \Delta S + w$$

where $w = u - u^*$.
As unobserved heterogeneity is removed the estimator of the coefficient of ΔS in (2) should be unbiased, while that of *S* in (1) will be subject to omitted variable bias. It could be easily shown.
If one fits the regression

$$\hat{G} = b_1 + b_2 S$$

using OLS, then

$$b_2 = \frac{\sum (S_i - \bar{S})(G_i - \bar{G})}{\sum (S_i - \bar{S})^2} =$$

$$= \frac{\sum (S_i - \bar{S})((\beta_1 + \beta_2 S_i + \beta_3 E_i + u_i) - (\beta_1 + \beta_2 \bar{S} + \beta_3 \bar{E} + \bar{u}))}{\sum (S_i - \bar{S})^2} =$$

$$= \beta_2 + \beta_3 \frac{\sum (S_i - \bar{S})(E_i - \bar{E})}{\sum (S_i - \bar{S})^2} + \frac{\sum (S_i - \bar{S})(u_i - \bar{u})}{\sum (S_i - \bar{S})^2}$$

Taking expectations, and making use of the invitation to treat *S* and *E* as nonstochastic,

$$E(b_2) = \beta_2 + \beta_3 \frac{\sum (S_i - \bar{S})(E_i - \bar{E})}{\sum (S_i - \bar{S})^2} + \frac{\sum (S_i - \bar{S})E(u_i - \bar{u})}{\sum (S_i - \bar{S})^2} =$$
$$= \beta_2 + \beta_3 \frac{\sum (S_i - \bar{S})(E_i - \bar{E})}{\sum (S_i - \bar{S})^2}$$

Hence the estimator is biased unless S and E happen to be uncorrelated in the sample**.** As a consequence, the standard errors will be invalid**.**

Discussion (optional).

However:

— it is possible that the bias in (1) may be small. This would be the case if E were a relatively unimportant determinant of G or if its correlation with S were low**.**

— it is possible that the variance in ΔS is smaller than that of S**.** This would be the case if S were changing slowly in each country, or if the rate of change of S were similar in each country**.**

Thus there may be a trade-off between bias and variance and it is possible that the estimator of β_2 using specification (1) could actually be superior according to some criterion such as the mean square error**.** Give an extra mark to any candidate who points out that the inclusion of δ_1 in (2) will make the estimation of δ_2 even less efficient.

[5 marks]

МАРКИНГ: В общем, конечно, для 5 баллов здесь написан явный перебор (в основе лондонская задача, хоть и переделанная, так что все из лондонской схемы я оставил).

Что нужно проверять, так это общее понимание панельных данных, понимание того, что незнание данных о предпринимательстве – это как раз и есть конкретный пример ненаблюдаемой неоднородности. Очень полезно привести общий вид модели, и потом с ним оперировать – про это лондонцы как раз забыли. Нужно понимание того, что при взятии разности эта неоднородность исчезает. Нужно понимание того, что если игнорировать ненаблюдаемые факторы, то возникает обычное смещение по типу ошибки спецификации. А вот делать математический вывод смещения вовсе не обязательно, да и вряд ли студентов сюда понесет, ведь явного задания не было. Кому вдруг это взбреднется – дать приз.

Вся дискуссия в конце помечена как optional, она конечно, очень умная, но ее появление здесь ничем, кроме шила в одном месте у экзаменатора, не может быть объяснено – тоже издержки явно смещенной лондонской схемы.

А вот сами соображения, которые в ней высказываются, как раз очень полезны и могут быть приведены почти в любом месте ответа студента – таким чудикам полный респект и дополнительный балл.

Особенно хорошо и уместно обсуждение «странности» уравнения – появления в нем константы. Студент ДОЛЖЕН понимать, что при дифференсинге никакие константы не выживают. Она там чужда, лишняя, снижает эффективности и прочая. Но для этого будет специальный пункт b).

(b) Explain why in principle you would expect the estimate of δ_1 in (2) not to be significant. Suppose that nevertheless the researcher finds that the coefficient is significant. Give at least two possible explanations, not counting the random error. What alternative method of fixed effect panel model could be suggested instead of differencing? Explain briefly.

b) If specification (1) is correct, there should be no intercept in (2) and for this reason the estimate of the intercept should not be significantly different from zero**.**

If it is significant, this could have occurred as a matter of Type I error**.** Close to this is explanation that uses multicollinearity (in this case multicollinearity with a constant). Alternatively, it might indicate a shift in the relationship between the two time periods**.** Suppose that (1) should have included a dummy variable set equal to 0 in the first time period and 1 in the second. d_1 would then be an estimate of its coefficient.

Some examples of such shift in relationship are:

1) If initial model contain time trend ($\ldots + kt$) after differencing we will get $d_1 = k$.
2) If unobserved variable E changed during the period under consideration.

There are at least two alternative methods: within group (WG) and LSDV methods. In WG we evaluate averages $\bar{G} = \frac{G + G^*}{2}$ and $\bar{S} = \frac{S + S^*}{2}$ and then use them for taking deviations from them $\Delta G = G - \bar{G}$,

$\Delta S = S - \overline{S}$ instead of direct differences $\Delta G = G - G^*$ and $\Delta S = S - S^*$. As data comprise only two periods there is not much difference between these two methods. In LSDV method we introduce 28 dummies D_i - one for each country (only 27 if we want to keep constant) and consider a model $G_{it} = \beta_2 S_{it} + \sum_{i=1}^{28} \gamma_i D_i + u_{it}$.
Additional variables allow to take under control unobserved factors via evaluation so called fixed effects (sixed differences in level between countries under consideration on the base of coefficients β_2 and γ_i.
[5 marks]
МАРКИНГ: Можно перебор в баллах в пункте а) паче чаяния оный случится, частично перенаправлять сюда, особенно в части «странной» константы. Но вообще-то и тут есть о чем поговорить.
Перечисляя пункты, которые стоит здесь проверить, сначала повторюсь: студент должен понимать, что при взятии разности константа должна сдохнуть, так что вроде как появилась она здесь, скорее всего, сдуру, из-за ошибочной спецификации. Но, оказывается, иногда полезно делать дурацкие ошибки. Значимость константы сразу должна настроить студента на серьезный лад – дыма без огня не бывает.
Дима Малахов здесь напомнил о возможности мультиколинеарности с константой. А настоящий экономист Замков предложил два возможных очень глубоких объяснения: скрытый неучтенный временной тренд, и сдвиг ненаблюдаемой переменной.
Проще всего здесь считать мысли по пальцам, студенты, вероятно, тоже будут использовать свои пальцы, чтобы высосать из них хоть что-то путное. Думаю, что на всю эту дискуссию достаточно будет пары баллов (плюс может призовой).
Остальные три я бы дал на вполне стандартный вопрос о возможных альтернативах методы разностей. Изложение вполне стандартное, но студент должен все изложить достаточно предметно, лучше с формулами, чтобы было ясно, что он понимает, о чем говорит. Например, только при детальном изложении понятно, что здесь по сути никакой разницы между методами разностей и WG тут нет (на это мы давали математическую задачку в прошлогоднем экзамене). LSDV должен быть изложен пристойно.

(c) Describe the advantages and disadvantages of random effects regressions compared with fixed effect regressions. Which test could help with this choice? Could the researcher have used a random effects regression in the present case?

<u>Solution and marking</u>
c) In fixed effects regressions (WG and FD), any regressor that is unvarying across time for each individual is washed out, or (in LSDV) we have to introduce a lot of dummies and so to reduce essentially the degrees of freedom (the same is in WG and FD). Random effects regressions do not suffer from this problems**.** They also preserve a greater number of degrees of freedom. The unobserved heterogeneity is simply added to disturbance term $G_{it} = \beta_1 + \beta_2 S_{it} + (\alpha_i + u_{it})$, based on the assumptions that α_i is generated by a ramdom process independently of S_{it}. To discriminate between fixed and ramdom effects we can use Darbin-Wu-Hausman test. This test uses chi-square statistic, based on comparison of estimated coefficients of fixed panel and random panel regressions, here with one degree of freedom (number of regressors). Under H_0 (which means that α_i are not correlated with S_j for any *i,j*) both fixed effect and random effect models provide us with consistent estimates. However fixed effect model estimates will be inefficient since it involves estimating an unnecessary set of coefficients, so random effect model should be used if H_0 is not rejected. But here this test is of no use as random effects requires the sample to be drawn randomly from a population and for unobserved effects to be uncorrelated with the regressors**.** The first condition is not satisfied here (European Union countries cannot be considered as sampled randomly), so random effects would be inappropriate here**.**
[5 marks]
МАРКИНГ: В общем, вопрос вполне теоретический и просто собирает все основное из панелей, что не вошло ранее в другие вопросы. Я вставил точные теоретические формулировки и рассуждения с большим запасом, поскольку совершено не ясно, куда и на что занесет студентов. Разумеется, все это здесь писать не нужно, но наговорить на 5 баллов не мешает.

5. The following ordinary least squares estimates were made on a sample of 3,356 employed male workers in Britain.

	total sample (i)	(ii)	under 40 (iii)	40 and over (iv)
ed	0.333 (0.047)	—	0.208 (0.067)	0.386 (0.067)
ed2	−0.159 (0.007)	—	−0.553 (0.036)	−1.087 (0.228)
age	0.142 (0.006)	0.151 (0.006)	0.262 (0.021)	0.128 (0.036)
age2	−0.159 (0.007)	−0.175 (0.007)	−0.362 (0.036)	−0.138 (0.035)
constant	0.002 (0.357)	2.647 (0.110)	−0.705 (0.558)	−0.277 (1.016)
n	3,356	3,356	1,807	1,549
R^2	0.233	0.180	0.312	0.135
RSS	1,235.88	1,610.19	607.89	608.73

The dependent variable is the log of weekly earnings. The estimates in the first two columns use the whole sample. The estimates in the third column use only observations on men younger than 40, the estimates in the fourth column use only observations on men aged 40 and over; *ed* is years of full time education, *ed2* is $ed \cdot ed/100$, *age* is age in years; *age2* is $age \cdot age/100$. n is the sample size. *RSS* is the residual sum of squares. Standard errors are given in brackets.

(a) Derive ordinary least squares estimators for the model specification (ii) of the type: $Y_i = \beta_0 + \beta_1 X_i + \beta_2 \dfrac{X^2{}_i}{100} + u_i$, where $Y_i = earning_i$; $X_i = age_i$ under appropriate assumptions on u_i and X_i.

Solution: OLS estimators minimise the residual sum of squares (*RSS*). Let $e_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i - \hat{\beta}_2 \dfrac{X^2{}_i}{100}$ be the residual, then $\sum e_i^2 = \sum \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i - \hat{\beta}_2 \dfrac{X^2{}_i}{100} \right)^2$ is the *RSS*. Minimising *RSS* with respect to $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$, we get the first-order conditions $\dfrac{\partial \sum e_i^2}{\partial \hat{\beta}_0} = 0$; $\dfrac{\partial \sum e_i^2}{\partial \hat{\beta}_1} = 0$ and $\dfrac{\partial \sum e_i^2}{\partial \hat{\beta}_2} = 0$.

$$\sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i + 0.01\hat{\beta}_2 \sum X_i^2$$

$$\sum Y_i X_i = \hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2 + 0.01\hat{\beta}_2 \sum X_i^3$$

$$\sum Y_i X_i^2 = \hat{\beta}_0 \sum X_i^2 + \hat{\beta}_1 \sum X_i^3 + 0.01\hat{\beta}_2 \sum X_i^4$$

Solving these first-order conditions we obtain the OLS estimators $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$.

Alternatively we can use well-known formulas for the OLS estimators of coefficients of multiple regression with two regressors

$$b_2 = \frac{\mathrm{Cov}(X_2,Y)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3,Y)\mathrm{Cov}(X_2,X_3)}{\mathrm{Var}(X_2)\mathrm{Var}(X_3) - [\mathrm{Cov}(X_2,X_3)]^2}$$

$$b_3 = \frac{\mathrm{Cov}(X_3,Y)\mathrm{Var}(X_2) - \mathrm{Cov}(X_2,Y)\mathrm{Cov}(X_2,X_3)}{\mathrm{Var}(X_2)\mathrm{Var}(X_3) - [\mathrm{Cov}(X_2,X_3)]^2}$$

$$b_1 = \bar{Y} - b_2\bar{X}_2 - b_3\bar{X}_3$$

So we get

$$b_2 = \frac{\text{Cov}(X,Y)\text{Var}(X^2) - \text{Cov}(X^2,Y)\text{Cov}(,X^2)}{\text{Var}(X)\text{Var}(X^2) - \left[\text{Cov}(X,X^2)\right]^2}$$

$$b_3 = \frac{\text{Cov}(X^2,Y)\text{Var}(X) - \text{Cov}(X,Y)\text{Cov}(X,X^2)}{\text{Var}(X)\text{Var}(X^2) - \left[\text{Cov}(X,X^2)\right]^2}$$

$$b_1 = \overline{Y} - b_2\overline{X} - b_3\overline{X^2}$$

(5 marks)

МАРКИНГ: Тут для студента главное – вовремя остановиться и не погрешить против стиля блистательной неопределенности, которым как раз отличаются лучшие ответы. Явные формулы с суммами никому не нужны. Готовые формулы ничуть не хуже. Важно, что студент что-то помнит и умеет делать. Разумеется, в хорошем стиле.

(b) Discuss the role of the quadratic terms in the models estimated above, economically, mathematically and graphically. Discuss the significance of the corresponding coefficients. Test the hypothesis that the coefficients of *ed* and *ed2* are jointly zero in full sample.

Solution: The quadratic terms show that in every case both education and age have a positive effect on earnings (the positive estimated coefficients of the linear terms) the rate of increase in earnings declines as age and education increase (the negative quadratic terms). Note all the coefficients (both linear and quadratic) of education and age are significant. So there is some optimal level of education as well as optimal age when

We test

$$H_0: \text{Coefficients are jointly equal to 0}$$

$$H_1: \text{Not all coefficients are equal to 0.}$$

Then

$$F = \frac{(1610.19 - 1235.88)/2}{1235.88/3351} = \frac{187.155}{0.3688} = 507.47\,.$$

Since the 5% critical value is $F_{2,\,3351,\,0.05} = 2.996$ (approximately), we reject H_0.

(5 marks)

МАРКИНГ: В условии сказано, что нужно делать и пояснение (экономическое или математическое) и график – пусть хоть параболу нарисуют. За все это вместе два балла. Если студент понимает, что на самом деле тут кривая поверхность в трехмерном пространстве, и что для каждого возраста существует свой оптимальный возраст (и наоборот), то это заслуживает призового балла (в пределах общих пяти баллов).
За внимание к значимости коэффициентов в любой форме еще балл.
И последние два балла за F-тест.

(c) Test the hypothesis that the coefficients of the model are constant between men under 40 and men aged 40 and over.

Solution: This is Chow test for the homogeneity of the sample

$$H_0: \text{The coefficients are constant}$$

$$H_1: \text{The coefficients are not constant.}$$

Then

$$F = \frac{(1235.88 - (607.89 + 608.73))/5}{(607.89 + 608.73)/3346} = \frac{3.852}{0.3636} = 10.59\,.$$

Since the 5% critical value is $F_{5,\,3346,\,0.05} = 2.214$ (approximately), we reject H_0.

(5 marks)

МАРКИНГ: В общем, тут все ясно. Тест или сделан или не сделан. Или с ошибками или без. Просто проверять все элементы теста.

(d) The alternative way to test the hypothesis in **c)** is to use dummy variable approach. Describe this approach in details using the model (i) (introduction of dummies, appropriate tests and so on). What are the relations between two approaches in **c)** and **d)**?

Solution: Consider theoretical equation corresponding to equation (1)

$$earning_i = \beta_0 + \beta_1 ed_i + \beta_2 ed2_i + \beta_3 age_i + \beta_4 age2_i + u_i\,;\ i = 1, 2, ..., n\,, \quad (1)$$

To take into account age categories we divide all men under consideration in our sample into two groups: under 40, and 40 and over, and introduce dummy variable D_i which is equal to 0 for all men from the first group (under 40) and is equal 1 for the men from the second group.

The simplest model with this dummy is

$$earning_i = \beta_0 + \gamma_0 D_i + \beta_1 ed_i + \beta_2 ed2_i + \beta_3 age_i + \beta_4 age2_i + u_i$$

But this model assumes that all coefficients β_1, β_2, β_3, β_4 are equal for two groups. To allow them to be different we have to define additionally four slope dummies $D_i \cdot ed_i$, $D_i \cdot ed2_i$, $D_i \cdot age_i$, $D_i \cdot age2_i$ and run equation

$$earning_i = \beta_0 + \gamma_0 D_i + \beta_1 ed_i + D_i \cdot ed_i + \beta_2 ed2_i + D_i \cdot ed2_i + \beta_3 age_i + D_i \cdot age_i + \beta_4 age2_i + D_i \cdot age2_i + u_i \quad (2)$$

to find sum of squared residuals RSS_U. We compare it with the RSS_R for the restricted model (1).

F-test, based on F-statistic $F = \dfrac{(RSS_R - RSS_U)/5}{RSS_U / 3346}$ is equivalent to Chow test in **(c)**.

(5 marks)

МАРКИНГ: Студент обязательно должен четко объяснить как оно вводит дамми. Выписывать модель с одним дамми не обязательно. До сих пор студент зарабатывает 1 балл. Остальные баллы даются за введение дамми наклона (2 балла), и за тест (1 балл). Последний балл студент зарабатывает за эквивалентность тестов.

6. Consider the model $y_t = \alpha x_t + u_t$; $t = 1, 2, ..., T$, where $\mathrm{E}(u_t) = 0$; $\mathrm{E}(u_t^2) = \sigma^2 x_t^2$; $\mathrm{E}(u_s u_t) = 0$ if $s \neq t$, for all s and t. x_t is an observed non-random variable. The density function of u_t is $f(u_t) = (2\pi\sigma^2 x_t^2)^{-1/2} \exp\left[-\frac{1}{2}\left(\frac{u_t}{\sigma x_t}\right)^2\right]$.

(a) Derive log-likelihood function for the model above as a function of parameters of α and σ^2. **(5 marks)**

Solution:

(a) From $y_t = \alpha x_t + u_t$ we get $u_t = y_t - x_t\alpha$ and so $\frac{u_t}{x_t} = \frac{y_t}{x_t} - \alpha$. Using expression for the density function $f(u_t) = (2\pi\sigma^2 x_t^2)^{-1/2} \exp\left[-\frac{1}{2}\left(\frac{u_t}{\sigma x_t}\right)^2\right]$ we evaluate the likelihood function as

$$L(u_1, ..., u_T; \alpha, \sigma^2) = \prod_{t=1}^{T} f(u_t) = \prod_{t=1}^{T} (2\pi\sigma^2 x_t^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}\left(\frac{u_t}{x_t}\right)^2\right] = \prod_{t=1}^{T} (2\pi\sigma^2 x_t^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}\left(\frac{y_t}{x_t} - \alpha\right)^2\right]$$

So the log-likelihood function is

$$\ln L = -\frac{T}{2}\log 2\pi - \frac{1}{2}\log \sum \sigma^2 x_t^2 - \frac{1}{2\sigma^2}\sum\left(\frac{y_t}{x_t} - \alpha\right)^2.$$

(5 marks)

МАРКИНГ: Я выделил этот пункт из последующего, так как на лондонском экзамене часто просят выписать функцию правдоподобия, а она здесь нетривиальна.

(b) Derive the maximum likelihood (ML) estimators of α and σ^2. Show that $\hat{\alpha}$ the ML estimator of α is unbiased.

Solution:

(b) The first-order conditions are:

$$\frac{\partial \ln L}{\partial \alpha} = -\frac{1}{2\sigma^2}\sum\left(\frac{y_t}{x_t} - \alpha\right) = 0 \qquad \text{(i)}$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4}\sum\left(\frac{y_t}{x_t} - \alpha\right)^2 = 0. \qquad \text{(ii)}$$

Solving (i) and (ii), the ML estimators of α and σ^2 are obtained as:

$$\hat{\alpha}_{MLE} = \frac{1}{T}\sum\left(\frac{y_t}{x_t}\right)$$

and

$$\hat{\alpha}^2_{MLE} = \frac{1}{T}\sum\left(\frac{y_t}{x_t} - \alpha\right)^2.$$

To show $\hat{\alpha}_{MLE}$ is an unbiased estimator,

$$\mathrm{E}(\hat{\alpha}_{MLE}) = \frac{1}{T}\sum\left(\frac{\mathrm{E}(y_t)}{x_t}\right) = \frac{1}{T}\sum\left(\frac{\alpha x_t}{x_t}\right) = \frac{T\alpha}{T} = \alpha$$

Hence $\hat{\alpha}_{MLE}$ is an unbiased estimator of α. **(5 marks)**

МАРКИНГ: Если предыдущий пункт выполнен правильно, то здесь работа чисто техническая.

(c) Compare the *ML* estimator of α , with the weighted least squares estimator of α .

Solution:
(a.ii) To obtain the weighted least squares (WLS) estimator, the model is divided by x_t to get

$$\frac{y_t}{x_t} = \alpha + \frac{u_t}{x_t}\,;\; u_t = y_t - \alpha x_t\,.$$

To obtain the WLS estimator of α ,

$$\sum\left(\frac{u_t}{x_t}\right)^2 = \sum\left(\frac{y_t}{x_t} - \alpha\right)^2$$

has to be minimised.
To obtain the *ML* estimator, In *L* is maximised, which is equivalent to minimizing

$$\sum\left(\frac{y_t}{x_t} - \alpha\right)^2.$$

Hence WLS estimators and *ML* estimators are the same.
(5 marks)
МАРКИНГ: Этот пункт можно выполнять даже при несделанном предыдущем.

(d) Consider a model $Y_t = \alpha + \beta X_t + u_t\,;\; t = 1, 2, ..., T$, where u_t is normally distributed with mean 0 and variance σ^2 and $\mathrm{E}(u_s u_t) = 0$ for all $s \neq t$. α and β have been estimated by maximum likelihood. Explain how the hypothesis that the coefficients are jointly equal to zero will be tested.

Solution: All the slope coefficients are equal to zero and can be tested using the likelihood ratio statistic $2(\log L - \log L_0)$. This is asymptotically distributed as a chi-square with $k-1$ degrees of freedom. $k-1$ is the number of explanatory variables in the model. In this case $k-1=2$.
$\log L$ is the log of the unrestricted likelihood and $\log L_0$ is the log of restricted likelihood.
$\log L_0$ has been obtained with only the intercept in the regression.
Pseudo- R^2 should also be discussed.
(5 marks)
МАРКИНГ: Совершенно стандартный вопрос, стоящий особняком.

7. In a certain bond market the demand for bonds, B_t, in period *t* is negatively related to the expected interest rate, i^e_{t+1}, in period $t+1$:

$$B_t = \beta_1 + \beta_2 i^e_{t+1} + u_t \qquad (1)$$

where u_t is a disturbance term not subject to autocorrelation. The expected interest rate is determined by an adaptive expectations process:

$$i^e_{t+1} - i^e_t = \lambda(i_t - i^e_t) \qquad (2)$$

where i_t is the actual rate of interest in period *t*. A researcher uses the following model to fit the relationship:

$$B_t = \gamma_1 + \gamma_2 i_t + \gamma_3 B_{t-1} + v_t \qquad (3)$$

where v_t is a disturbance term.

(a) Show how this model may be derived from the demand function and the adaptive expectations process.

Solution and marking

a) The adaptive expectations process may be rewritten

$$i^e_{t+1} = \lambda i_t + (1-\lambda) i^e_t$$

Substituting this into (1), one obtains

$$B_t = \beta_1 + \beta_2 \lambda i_t + \beta_2 (1-\lambda) i^e_t + u_t$$

We note that if we lag (1) by one time period,

$$B_{t-1} = \beta_1 + \beta_2 i^e_t + u_{t-1}$$

Hence

$$\beta_2 i^e_t = B_{t-1} - \beta_1 - u_{t-1}$$

Substituting this into the second equation above, one has

$$B_t = \beta_1 \lambda + \beta_2 \lambda i_t + (1-\lambda) B_{t-1} + u_t - (1-\lambda) u_{t-1}$$

This is equation (3) in the question, with $\gamma_1 = \beta_1 \lambda$, $\gamma_2 = \beta_2 \lambda$, $\gamma_3 = 1-\lambda$ and $v_t = u_t - (1-\lambda) u_{t-1}$.

[5 marks]

МАРКИНГ: Совершенно классическая схема вывод модели ADL из схемы адаптивных ожиданий преобразованием Койка, то есть сразу, в один ход. Оно весьма изобретательно, и если не помнить его схемы, то вполне можно и не догадаться до хитрой подстановки.

Можно предположить, что значительная часть студентов пустится во все тяжкие, повторяя лекционную длиннющую схему с бесконечной подстановкой и получения бесконечного ряда. Поскольку это не является ответом на вопрос, то ВСЕ НАБРАННЫЕ ЗДЕСЬ БАЛЛЫ РАЗУМНО ПЕРЕНЕСТИ В ПУНКТ с), где прямо просят вывести бесконечный ряд Койка (разумеется, в пределах максимума 5 баллов за пункт). Тогда, понятно, никаких баллов за пункт b) студент не получает.

(b) Explain why inconsistent estimates of the parameters will be obtained if equation (3) is fitted using ordinary least squares (OLS). (A mathematical proof is not required. Do not attempt to derive an expression for the bias.)

Solution and marking

b) In equation (3), the regressor B_{t-1} is partly determined by u_{t-1}. The disturbance term v_y also has a component u_{t-1}. Hence the requirement that the regressors and the disturbance term be distributed independently of each other is violated. The violation will lead to inconsistent estimates because the regressor and the disturbance term are contemporaneously correlated.

[5 marks]

МАРКИНГ: Это премиальный вопрос для тех, кто выучил одноходовое преобразование Койка. Только проследить, что студент четко понимает что там с чем коррелирует. Общих слов недостаточно.

(c) Describe a method for fitting the model that would yield consistent estimates.

Solution and marking

c) If the first equation in (a) above is true for time period $t+1$, it is true for time period *t*:

$$i_t^e = \lambda i_{t-1} + (1-\lambda) i_{t-1}^e$$

Substituting into the second equation in (a), we now have

$$B_t = \beta_1 + \beta_2 \lambda i_t + \beta_2 (1-\lambda) i_{t-1} + (1-\lambda)^2 i_{t-1}^e + u_t$$

Continuing to lag and substitute, we have

$$B_t = \beta_1 + \beta_2 \lambda i_t + \beta_2 \lambda (1-\lambda) i_{t-1} + \ldots + \beta_2 \lambda (1-\lambda)^{s-1} i_{t-s+1} + (1-\lambda)^s i_{t-s+1}^e + u_t$$

For *s* large enough, $(1-\lambda)^s$ will be so small that we can drop the unobservable term i_{t-s+1}^e with negligible omitted variable bias. The disturbance term is distributed independently of the regressors and hence we obtain consistent estimates of the parameters**.** The model should be fitted using a nonlinear estimation technique that takes account of the restrictions implicit in the specification.

[5 marks]

МАРКИНГ: Здесь приведена очень компактная форма вывода. Студенты развезут на страницу. Но недостаточно просто повозиться с формулами. Нужно обсудить все аспекты проблемы оценивания: наличие пренебрежимо малого смещения, отсутствие связи случайного члена с регрессорами, и получение в результате состоятельных оценок. Важно упоминание НЕЛИНЕЙНОГО OLS. Вполне допустим также рассказ о том, что можно реализовать этот нелинейный метод вручную путем поиска наилучшего в смысле RSS значения «лямбда» - сначала фиксируется значение, близкое к нулю, вычисляется вся правая часть, гонится регрессия левой части по правой (тут вообще никаких проблем) и вычисляется RSS. Потом пробуется другое значение «лямбда» чуть ближе к единице и т.д.

В общем, тут есть место подвигу.

(d) Suppose that u_t were subject to the first-order autoregressive process: $u_t = \rho u_{t-1} + \varepsilon_t$, where ε_t is not subject to autocorrelation. How would this affect your answer to (b)?

ВСЕ ПРАВИЛЬНО – ИМЕННО b)

Solution and marking

d) v_t is now given by

$$v_t = u_t - (1-\lambda) u_{t-1} = \rho u_{t-1} + \varepsilon_t - (1-\lambda) u_{t-1} = \varepsilon_t - (1-\rho-\lambda) u_{t-1}.$$

Since ρ and λ may reasonably be assumed to lie between 0 and 1, it is possible that their sum is approximately equal to 1, in which case v_t is approximately equal to the innovation ε_t. If this is the case, there would be no violation of the regression assumption described in (b) and one could use OLS to fit (3) after all.

[5 marks]

МАРКИНГ: Это просто кусок из учебника (в последнем издании приведен для большей заметности на врезке), где показано, что в эконометрике часто клин клином вышибают. Как и в жизни, два недостатка иногда превращаются в большое достоинство, и все становится тип-топ. Понятно, что сюда доберутся только те, кто что-то читал.

8. Consider the following processes:

$k_t = -k_{t-1} + \varepsilon_t$,

$y_t = y_{t-1} + u_t$,

$p_t = y_t^2 + v_t$,

$r_t = 2 + 3t + \eta_t$,

$x_t = 5 + 8t + w_t$,

$z_t = 2y_{t-1} + s_t$,

$h_t = 0.5h_{t-1} + \mu_t$.

Assume, that ε_t, u_t, v_t, η_t, w_t, s_t, μ_t are iid $N(0,\sigma^2)$ and are independently distributed from each other and variables k_t, y_t, p_t, r_t, x_t, z_t, h_t.

(a) Define for each process whether it is stationary or not. For nonstationary processes define whether they are Difference Stationary (DS) or Trend Stationary (TS) or another type of nonstationary processes.

Solution.

$k_t = -k_{t-1} + \varepsilon_t$, (another),

$y_t = y_{t-1} + u_t$, (DS),

$p_t = y_t^2 + u_t$, (another),

$r_t = 2 + 3t + \eta_t$, (TS),

$x_t = 5 + 8t + w_t$, (TS),

$z_t = 2y_{t-1} + s_t$, (another),

$h_t = 0.5h_{t-1} + \mu_t$, (stationary)/

(b) If pairwise cointegrated processes here exist, find them and corresponding cointegrating vector. Explain your solution.

There are one pair of cointegrated vectors: y_t, z_t. Cointegrating vector (up to constant multiplier) is $(2,-1)$

.

(c) Let's consider the following true economic model: $y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 z_{t-2} + \alpha_4 z_{t-3} + \tau_t$ with the variables y_t and z_t defined and discussed in **(a)** and **(b)**, where $\tau_t \sim iid$ with zero expectation and σ^2 variance. Find the long run relation of variables y_t and z_t.What is the difference between the short run and long run dynamics?

Solution.

$\bar{y} = \alpha_0 + \alpha_1 \bar{y} + \alpha_2 \bar{y} + \alpha_3 \bar{z} + \alpha_4 \bar{z}$. Thus, $\bar{y} = \dfrac{\alpha_0}{1-\alpha_1-\alpha_2} + \dfrac{\alpha_3+\alpha_4}{1-\alpha_1-\alpha_2}\bar{z}$.

(d) Show in detail how to estimate the model $y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \alpha_3 z_{t-2} + \alpha_4 z_{t-3} + \tau_t$ from **(c)** using error correction model.

Solution.
Note:

$$y_{t-2} = y_{t-1} - \Delta y_{t-1}$$

$$z_{t-2} = z_{t-1} - \Delta z_{t-1}$$

$$z_{t-3} = z_{t-2} - \Delta z_{t-2}$$

Thus:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 (y_{t-1} - \Delta y_{t-1}) + \alpha_3 (z_{t-1} - \Delta z_{t-1}) + \alpha_4 (z_{t-2} - \Delta z_{t-2}) + \tau_t$$

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-1} - \alpha_2 \Delta y_{t-1} + \alpha_3 z_{t-1} - \alpha_3 \Delta z_{t-1} + \alpha_4 z_{t-2} - \alpha_4 \Delta z_{t-2} + \tau_t$$

$$y_t = \alpha_0 + (\alpha_1 + \alpha_2) y_{t-1} + \alpha_3 z_{t-1} + \alpha_4 z_{t-2} - \alpha_2 \Delta y_{t-1} - \alpha_3 \Delta z_{t-1} - \alpha_4 \Delta z_{t-2} + \tau_t$$

$$y_t = \alpha_0 + (\alpha_1 + \alpha_2) y_{t-1} + \alpha_3 z_{t-1} + \alpha_4 (z_{t-1} - \Delta z_{t-1}) - \alpha_2 \Delta y_{t-1} - \alpha_3 \Delta z_{t-1} - \alpha_4 \Delta z_{t-2} + \tau_t$$

$$y_t = \alpha_0 + (\alpha_1 + \alpha_2) y_{t-1} + (\alpha_3 + \alpha_4) z_{t-1} - \alpha_2 \Delta y_{t-1} - (\alpha_3 + \alpha_4) \Delta z_{t-1} - \alpha_4 \Delta z_{t-2} + \tau_t$$

Subtract from both sides y_{t-1}:

$$\Delta y_t = \alpha_0 + (\alpha_1 + \alpha_2 - 1) y_{t-1} + (\alpha_3 + \alpha_4) z_{t-1} - \alpha_2 \Delta y_{t-1} - (\alpha_3 + \alpha_4) \Delta z_{t-1} - \alpha_4 \Delta z_{t-2} + \tau_t$$

$$\Delta y_t = (\alpha_1 + \alpha_2 - 1)\left(y_{t-1} - \frac{\alpha_0}{(1 - \alpha_1 - \alpha_2)} - \frac{(\alpha_3 + \alpha_4)}{(1 - \alpha_1 - \alpha_2)} z_{t-1} \right) - \alpha_2 \Delta y_{t-1} - (\alpha_3 + \alpha_4) \Delta z_{t-1} - \alpha_4 \Delta z_{t-2} + \tau_t$$

We can estimate

$y_{t-1} - \dfrac{\alpha_0}{(1 - \alpha_1 - \alpha_2)} - \dfrac{(\alpha_3 + \alpha_4)}{(1 - \alpha_1 - \alpha_2)} z_{t-1}$ as a residuals, e_{t-1}, from model $y_{t-1} = \gamma_0 + \gamma_1 z_{t-1} + \zeta_t$.

Because y_t, z_t are cointegrated, than e_{t-1} is stationary. Δy_t and Δz_t are stationary because $y_t \in I(1)$, $z_t \in I(1)$. So we can use OLS to estimate this model.

The International College of Economics and Finance
Econometrics – 2016-2017.
Midterm exam 2017. March 30.
Suggested Solutuins

General instructions. Candidates should answer SIX of the following SEVEN questions: all 4 questions of the Section A and any 2 of the questions from Section B (questions 5-8). The weight of the Section A is 60% of the exam; three other questions from the Section B add 20% each. You are advised to divide your time accordingly. Structure your answers in accordance with the structure of the questions. When testing hypotheses always state clearly null and alternative hypotheses provide critical value used for test, mentioning degrees of freedom and the significance level chosen for the test.

SECTION A
(1 hour 50 minutes)
Answer **ALL** questions from this section (questions **1-4**).
Each question in this section bears 15 marks

1. (a) A student has a representative sample of 25 residents of a small town in United States. The relationship between the resident' expenditure on potato chips $CHIPS_i$ (in of dollars) and personal disposable income DPI_i (also in of dollars) is studied, this relationship is described with the following equation:

$$CHIPS_i = 5.3063 - 0.27683 \log DPI_i$$
$$(s.e.) \quad (1.0242) \; (0.15445)$$

(a) Perform t-tests for significance of both coefficients of the model and give interpretation to them.

Solution
First interpret the coefficients. The slope (if it turned out to be significant) could be interpreted as decrease by 0.27 cents ($\frac{-0.27683}{100} = -0.0027683$ dollars) of the mean expenditures on chips consumption caused by 1% increase of income. Intercept seems to have no meaningful interpretaion.
Consider theoretical equation corresponding estimated equation under consideration:

$$CHIPS_i = \beta_1 + \beta_2 \log DPI_i + u_i$$

Under the pair of hypotheses $H_0 : \beta = 0$; $H_a : \beta \neq 0$ the critical values for t-test are $t_{crit}^{5\%} = 2.069$ and $t_{crit}^{1\%} = 2.807$ for $25 - 2 = 23$ degrees of freedom. The values of t-statistic are $t = \frac{-0.27683}{0.15445} = -1.7924$ for the slope and $t = \frac{5.3063}{1.0242} = 5.18$ for the intercept, so the slope is insignificant ($|-1.7924| < 2.069$) while the intercept is significant ($5.18 > 2.807$). So the hypothesis $H_0 : \beta_2 = 0$ is not rejected (income has no significant influence on the expenditures on chips consumption), while $H_0 : \beta_1 = 0$ is rejected, so equation is in fact $CHIPS_i = \beta_1 + 0 \cdot \log DPI_i + u_i = \beta_1 + u_i$. So the value of the intercept equal to $\$5.3063$ could be interpreted as the estimate of the mean expenditures of residents on chips consumption.
[5 marks]

(b) At the seminar one of the participants remarked that coefficient of log(DPI) is expected to be negative as people tend to buy healthy food instead of chips as their income rises. The other participant objected to him saying that just the opposite is true (the coefficient of log(DPI) should be positive) as chips consumption is used mainly by young people with low income who cannot buy more expensive healthy food and are quite sensitive to changes in income. How both suggestions change your conclusion on significance of the coefficients of the model?

Solution
Let the equation above corresponds to the model $CHIPS_i = \beta_1 + \beta_2 \log(DPI_i) + u_i$. Then assuming the first assumption (the influence of the income on the expenditures on chips consumption cannot be positive) is true

one can use one sided test with $H_0:\beta_2=0$ (or $H_0:\beta_2\geq 0$) against $H_a:\beta_2<0$. Now $t_{crit}^{5\%}(one\ sided)=1.714$ and $-1.7924<-1.714$ so the coefficient β_2 of $\log(DPI_i)$ becomes significant.

The opposite assumption $H_0:\beta_2=0$ (or $H_0:\beta_2\leq 0$) against $H_a:\beta_2>0$ cannot change the conclusion of insignificance of the slope as actual value of $t=-1.7924$ is negative.

[5 marks]

(c) The third participant remarked that there is no sense at all in talking about this equation as R^2 is unknown. Help the researcher to recover the value of R^2, and based on it to perform F-test for the significance of the equation as a whole. Is this model suitable for predicting the consumption of chips? Compare your results in **(b)** and **(c)**: do they contradict each other? Explain.

Solution

R^2 could be easely restored from the knowledge of the *t*-statistics. As $t^2=(-1.7924)^2=3.2127=F$ so solving equation $F=\dfrac{R^2/1}{(1-R^2)/23}=3.2127$ for R^2 we get $R^2=0.1226$, so only about 12% of the variance in expenditures on chips consumption is explained by the changes in logarithm of income. Any prediction based on the model with such a low R^2 is senseless but any analytical work would be possible in the case of the significance of the slope.

In our case F-test shows that the equation is insignificant. This conclusion follows also from two sided t-test for the slope as these two tests are equivalent

There is no contradiction here as F-test is equivalent to the two sided t-test, while one sided test showed the significance of the slope under one the assumptions.

[5 marks]

Question 2.

2. Answer the following questions:

(a) Explain what a trend stationary series and what a difference stationary series are. What is an important difference between the two types of stationarity? If time series is difference stationary how it can be transformed into a stationary one? If time trend is detected how to get rid of it?

Solution: In part (a) the meaning of a trend stationary series and a difference stationary series should be explained. Part (b) requires a discussion of the Dickey-Fuller test, and in part (c) the error correction model (ECM) should be derived and the interpretation of the coefficients of the ECM should be given. The answer is as follows.

(a) If after removing the trend from a nonstationary series (detrending) the resulting variable becomes stationary, then the variable is called *trend stationary*. Let:

$$Z_t=X_t-\alpha_1 t=\alpha_0+u_t,$$

where $\mathrm{E}(u_t)=0$, $\mathrm{var}(u_t)=\sigma^2$ and $\mathrm{E}(u_t u_{t-s})=0$ for all *s* and *t*, then:

$$\mathrm{E}(Z_t)=\mathrm{E}(\alpha_0+u_t)=\alpha_0,$$

$$\mathrm{var}(Z_t)=\mathrm{Var}(\alpha_0+u_t)=\sigma^2,$$

$$\mathrm{cov}(Z_t,Z_{t-s})=\mathrm{E}[(Z_t-\mathrm{E}(Z_t))(Z_{t-s}-\mathrm{E}(Z_{t-s}))]=\mathrm{E}(u_t u_{t-s})=0.$$

This means that Z_t has constant mean and variance for all *t*, and covariance is zero for all $s > 0$. It implies that the series is trend stationary.
Alternatively for practical purposes the time variable could be included into the regression model.

If a nonstationary process can be transformed into a stationary process by differencing then the series is said to be *difference stationary*.
Let X_t be a random walk with a drift:

$$X_t = \beta_0 + X_{t-1} + \varepsilon_t, \qquad \text{(i)}$$

where $\mathrm{E}(\varepsilon_t) = 0$, $\mathrm{var}(\varepsilon_t) = \sigma^2$ and $\mathrm{E}(\varepsilon_t \varepsilon_s) = 0$ for all *s* and *t*, $s \neq t$.
Subtract X_{t-1} from both sides of (i) to get:

$$\Delta X_t = X_t - X_{t-1} = \beta_0 + \varepsilon_t.$$

It can be easily checked that $\mathrm{E}(\Delta X_t) = \beta_0$, $\mathrm{var}(\Delta X_t) = \sigma_\varepsilon^2$ and $\mathrm{cov}(\Delta X_t, \Delta X_{t-s}) = 0$ for all *s* and *t*, $s \neq t$. This means that ΔX_t is stationary. This implies that X_t is difference stationary.

It is interesting to note that trend stationary time series instead of detrending can be differenced.
If
$X_t = \alpha_0 + \alpha_1 t + u_t$
then
$X_{t-1} = \alpha_0 + \alpha_1 (t-1) + u_{t-1}$
so subtracting equations we get stationary time series
$\Delta X_t = X_t - X_{t-1} = \alpha_1 + u_t - u_{t-1}$
Nevertheless it is called trend stationary time series.

Its disturbance term $\varepsilon_t = u_t - u_{t-1}$ follows MA(1) process and so possibly posess some (negative) autocorrelation properties.

The difference between two types of time series.
It is important to know whether a variable is difference stationary or trend stationary because for difference stationary variables shocks have a permanent effect whereas for trend stationary variables shocks are transitory.
[5 marks]

(b) Let the model be:

$$Y_t = \phi Y_{t-1} + u_t;\ t = 1, 2, ..., T,$$

where u_t are independently and identically distributed as $N(0, \sigma^2)$. Explain how would you test $H_0 : \phi = 1$ against the alternative of a zero mean, covariance stationary $AR(1)$ process. Give the assumptions this test requires.

Solution:
The standard test for a unit root is due to Dickey and Fuller. In order to test the null hypothesis of a random walk without drift against the alternative of a zero mean covariance stationary $AR(1)$ process:

$$y_t = \phi y_{t-1} + \varepsilon_t$$

subtract y_{t-1} from both sides:

$$y_t - y_{t-1} = \phi y_t - y_{t-1} + \varepsilon_t,$$
$$\Delta y_t = (\phi - 1) y_{t-1} + \varepsilon_t = \rho y_{t-1} + \varepsilon_t.$$

Test $H_0 : \rho = 0$ using:

$$\hat{\tau} = \frac{\hat{\rho} - 0}{\text{s.e.}(\hat{\rho})}.$$

We cannot use the standard *t* test procedure in this case because the distribution of the statistic is not a *t* distribution, so critical values have to be computed by Dickey and Fuller using Monte Carlo techniques (Dickey-Fuller tables). The test is sensitive to the presence of serial correlation in the error term so we need to take steps to remove the effects of this serial correlation – this is done by including lagged values of Δy_t in the regression. The statistic has the same asymptotic distribution as $\hat{\tau}$.

[5 marks]

(c) The researcher wants to test whether time series X_t is stationary using augmented Dickey-Fuller test without time trend including three additional lags: $X_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-4} + u_t$ where u_t is distributed independently of X_t with zero mean and constant variance.

Derive Dickey-Fuller equation corresponding time series model under consideration. What is the null hypothesis? What is the decision rule?

Solution: (a) Starting from the equation above $X_t = \beta_0 + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-4} + u_t$ we first subtract the same term from both sides:

$$X_t - X_{t-1} = \beta_0 + \beta_1 X_{t-1} - X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-4} + u_t$$

Then add to and simultaneously subtract from the right side of equation the same terms $\beta_2 X_{t-1}$, $\beta_3 X_{t-1}$ and $\beta_4 X_{t-1}$:

$$X_t - X_{t-1} = \beta_0 + \beta_1 X_{t-1} - X_{t-1} + \beta_2 X_{t-1} - \beta_2 X_{t-1} + \beta_3 X_{t-1} - \beta_3 X_{t-1} + \beta_4 X_{t-1} - \beta_4 X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + u_t$$

and rearrange terms taking out the common factors of the brackets

$$X_t - X_{t-1} = \beta_0 + (\beta_1 + \beta_2 + \beta_3 + \beta_4 - 1) X_{t-1} - (\beta_2 + \beta_3 + \beta_4) X_{t-1} + \beta_2 X_{t-2} + \beta_3 X_{t-3} + u_t$$

The coefficient of variable X_{t-1}: $\beta_1 + \beta_2 + \beta_3 + \beta_4 - 1 = \gamma$ is the most important for the test. But we need represent all term in the right side of the equation in difference form (that are stationary).

Now we need to find a pair with additional lag for the term $-(\beta_2 + \beta_3 + \beta_4) X_{t-1}$ so simultaneously add to and subtract from the right side of equation the same terms $(\beta_3 + \beta_4) X_{t-2}$.

$$\Delta X_t = \beta_0 + \gamma X_{t-1} - (\beta_2 + \beta_3 + \beta_4) X_{t-1} + \beta_2 X_{t-2} + (\beta_3 + \beta_3) X_{t-2} - (\beta_3 + \beta_3) X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-4} + u_t$$

Now we have first difference in the right side of equation

$$\Delta X_t = \beta_0 + \gamma X_{t-1} - (\beta_2 + \beta_3 + \beta_4)(X_{t-1} - X_{t-2}) - (\beta_3 + \beta_3) X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-4} + u_t$$

Now we need to find a pair for the term $-(\beta_3 + \beta_4) X_{t-2}$ so simultaneously add to and subtract from the right side of equation the same terms $\beta_4 X_{t-3}$

$$\Delta X_t = \beta_0 + \gamma X_{t-1} - (\beta_2 + \beta_3 + \beta_4) \Delta X_{t-1} - (\beta_3 + \beta_4) X_{t-2} + \beta_3 X_{t-3} + \beta_4 X_{t-4} - \beta_4 X_{t-4} + \beta_4 X_{t-4} + u_t$$

And finally rearrange terms

$$\Delta X_t = \beta_0 + \gamma X_{t-1} - (\beta_2 + \beta_3 + \beta_4) \Delta X_{t-1} - (\beta_3 + \beta_4) \Delta X_{t-2} - \beta_4 \Delta X_{t-3} + u_t$$

Null hypothesis here is $H_0: \gamma = \beta_1 + \beta_2 + \beta_3 + \beta_4 - 1 = 0$ against alternative $H_a: \gamma = \beta_1 + \beta_2 + \beta_3 + \beta_4 - 1 < 0$.

The decision rule is following: if H_0 is not rejected the series X_t is non stationary.

[5 marks]

Question 3.

3. Derive the order of integration of x_t in the following models. Assume in each case that u_t is stationary, initial states (e.g. x_0) are fixed and $\mathrm{E}(u_s u_t)=0$ if $s \neq t$. It is enough to investigate the series itself, and if necessary its first and second differences.

(a) $x_t = \alpha_0 + u_t + u_{t-1}$; $t = 1, 2, ..., T$.

Solution: In both (a) and (b), it should be shown that the mean and the variance of the variable is constant and the covariances are independent of time but may depend on the length of the lag. Then one should try to transform it and check whether transformed time series is stationary. The solution is as follows.

First, make a full list of assumptions related disturbance term u_t: as usual $\mathrm{E}\,u_t = 0$, $\mathrm{var}\,u_t = \sigma_u^2$, and as it is said above $\mathrm{E}(u_s u_t)=0$ if $s \neq t$.

(a)

The series x_t is (weakly) stationary if

1) $\mathrm{E}(x_t) = const.$, that is $\mathrm{E}(x_t)$ does not depend on time;

2) $\mathrm{var}(x_t) = const.$, that is $\mathrm{var}(x_t)$ does not depend on time;

3) $\mathrm{cov}(x_t, x_{t-s}) = f(s)$, that is covariance of x_t and x_{t+s} depends only on the lag length s, and does not depend on time t.

Let us check all three conditions in turn.

$\mathrm{E}(x_t) = \mathrm{E}(\alpha_0 + u_t + u_{t-1}) = \alpha_0 + \mathrm{E}\,u_t + \mathrm{E}\,u_{t-1} = \alpha_0$, as $\mathrm{E}\,u_t = 0$

$\mathrm{var}(x_t) = \mathrm{var}(u_t - u_{t-1}) = \mathrm{var}(u_t) + \mathrm{var}(u_{t-1}) + 2\,\mathrm{cov}(u_t, u_{t-1}) = 2\sigma^2$ as $\mathrm{cov}(u_t, u_{t-1}) = \mathrm{E}(u_s u_t) = 0$ if $s \neq t$.

For covariance consider several cases.

$s=1$. Then $\mathrm{cov}(\alpha_0 + u_t + u_{t-1}, \alpha_0 + u_{t-1} + u_{t-2}) = \mathrm{cov}(u_t, u_{t-1}) + \mathrm{cov}(u_{t-1}, u_{t-2}) + \mathrm{cov}(u_{t-1}, u_{t-1}) + \mathrm{cov}(u_{t-1}, u_{t-2}) =$

$= \mathrm{E}(u_t u_{t-1}) + \mathrm{E}(u_{t-1} u_{t-2}) + \mathrm{E}(u_{t-1} u_{t-1}) + \mathrm{E}(u_{t-1} u_{t-2}) = 0 + 0 + \sigma_u^2 + 0 = \sigma_u^2$.

$s=2$. Then $\mathrm{cov}(\alpha_0 + u_t + u_{t-1}, \alpha_0 + u_{t-2} + u_{t-3}) = \mathrm{cov}(u_t, u_{t-2}) + \mathrm{cov}(u_{t-1}, u_{t-2}) + \mathrm{cov}(u_{t-1}, u_{t-3}) + \mathrm{cov}(u_{t-1}, u_{t-3}) =$

$= \mathrm{E}(u_t u_{t-2}) + \mathrm{E}(u_{t-1} u_{t-2}) + \mathrm{E}(u_{t-1} u_{t-2}) + \mathrm{E}(u_{t-1} u_{t-3}) = 0 + 0 + 0 + 0 = 0$.

The same result will be obtained in case $s > 2$.

So $\mathrm{cov}(x_t, x_{t-s}) = \sigma_u^2$ for $s = 1$ and $\mathrm{cov}(x_t, x_{t-s}) = 0$ for $s > 1$, and so it is function in s but not in t.

Therefore, x_t is stationary or $I(0)$.

[5 marks]

(b) $x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 t + u_t$; $|\alpha_1| < 1$; $t = 1, 2, ..., T$.

Solution:

As time series $x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 t + u_t$ contains time trend it is non stationary. To prove it let us take expectations: $\mathrm{E}(x_t) = \alpha_0 + \alpha_1\,\mathrm{E}(x_{t-1}) + \alpha_2 t$, so $\mathrm{E}(x_t)$ is clearly a function of time and $\mathrm{E}(x_t) \neq \mathrm{E}(x_{t-1})$.

Therefore, x_t is non-stationary.

Lag this equation by one period to get:

$$x_{t-1} = \alpha_0 + \alpha_1 x_{t-2} + \alpha_2 (t-1) + u_{t-1}.$$

Subtracting this from the first equation gives:

$$\Delta x_t = \alpha_1 \Delta x_{t-1} + \alpha_2 + u_t - u_{t-1}.$$

Since $|\alpha_1| < 1$, Δx_t follows a stationary $AR(1)$ process.

Hence x_t is $I(1)$ and it is trend stationary.

[5 marks]

(c) $x_t = x_{t-1} + x_{t-2} + u_t$

Solution: The simplest way to show that this time series is non stationary is to transform it into the form of Dickey-Fuller equation: subtracting $-x_{t-1}$ from the both side of equation.

$x_t = x_{t-1} + x_{t-2} + u_t$

we get

$x_t - x_{t-1} = x_{t-1} - x_{t-1} + x_{t-2} + u_t$ or $\Delta x_t = x_{t-1} - \Delta x_{t-1} + u_t$

This equation could be considered as the restricted version of the following Dickey-Fuller equation

$x_t - x_{t-1} = \beta_1 x_{t-1} - \beta_2 (x_{t-1} - x_{t-2}) + u_t$ or $\Delta x_t = \beta_1 x_{t-1} - \beta_2 \Delta x_{t-1} + u_t$

with the restrictions $H_0 : \beta_1 = 1,\ \beta_2 = 1$ against $H_a : \beta_1 < 1,\ \beta_2 \neq 1$. The restriction $\beta_1 = 1$ implies that x_t is not stationary.

It is obvious that neither $\Delta x_t = x_t - x_{t-1} = x_{t-2} + u_t$ nor $\Delta^2 x_t = \Delta x_t - \Delta x_{t-1}$ are not stationary as differenced time series $\Delta x_t = x_t - x_{t-1} = x_{t-2} + u_t$ is the lagged original time series again of the same order of interation. In fact $x_t \in I(\infty)$.

Alternative solution of the first part of the problem:

As time series (c) is likely to be non-stationary it is sufficient to find violation of some of conditions of stationarity. The solution is as follows.

In order to show that this time series is not stationary we also can transform it by sunsequent lagging and substitutung

$$x_{t-1} = x_{t-2} + x_{t-3} + u_{t-1},$$

and so

$$x_t = (x_{t-2} + x_{t-3} + u_{t-1}) + x_{t-2} + u_t = 2x_{t-2} + x_{t-3} + u_{t-1} + u_t.$$

Then

$$x_{t-2} = x_{t-3} + x_{t-4} + u_{t-2},$$

and so

$$x_t = 2(x_{t-3} + x_{t-4} + u_{t-2}) + x_{t-3} + u_{t-1} + u_t = 3x_{t-3} + 2x_{t-4} + 2u_{t-2} + u_{t-1} + u_t.$$

Once more

$$x_{t-3} = x_{t-4} + x_{t-5} + u_{t-3},$$

and so

$$x_t = 3(x_{t-4} + x_{t-5} + u_{t-3}) + 2x_{t-4} + 2u_{t-2} + u_{t-1} + u_t = 5x_{t-4} + 3x_{t-5} + 3u_{t-3} + 2u_{t-2} + u_{t-1} + u_t.$$

Now it is clear that

$$x_t = F_1 u_t + F_2 u_{t-1} + F_3 u_{t-2} + \ldots + F_{t-1} u_2 + F_t x_1 + F_{t+1} x_0.$$

where x_0, x_1 are two initial subsequent states of the stochastic process (which can be assumed as constant terms), and $F_1, F_2, \ldots$ are some integer coefficients (in could be clear seen that these coefficients are member of Fibonacci sequence 1, 1, 2, 3, 5, 8, …, with the following recurrence relation 2=1+1, 3=2+1, …., $F_n = F_{n-1} + F_{n-2}$ but this fact is not essential for our purpose, it is sufficient to note that all of them are natural numbers so $F_n \geq 1$).

Then

$$\mathrm{E}\, x_t = F_1 \mathrm{E}\, u_t + F_2 \mathrm{E}\, u_{t-1} + F_3 \mathrm{E}\, u_{t-2} + \ldots + F_{t-1} \mathrm{E}\, u_2 + F_t x_1 + F_{t+1} x_0 = F_t x_1 + F_{t+1} x_0 = const..$$

Now let us evaluate variance, omitting zero terms of the type $\mathrm{cov}(u_t, u_s); t \neq s$:

$\operatorname{var} x_t = (F_1)^2 \operatorname{var} u_t + (F_2)^2 \operatorname{var} u_{t-1} + ... + (F_{t-1})^2 \operatorname{var} u_2 = \sigma_u^2 (1^2 + 1^2 + 2^2 ... + (F_{t-1})^2) \to +\infty$ as $t \to +\infty$ (taking into account that all $F_n \geq 1$). So the second condition of stationarity $\operatorname{var} x_t = const$ is violated, ans so x_t is non-stationary.

Difference transformation does not hepl here: neither $\Delta x_t = x_t - x_{t-1} = x_{t-2} + u_t$ nor $\Delta^2 x_t = \Delta x_t - \Delta x_{t-1}$ are not stationary as differenced time series $\Delta x_t = x_t - x_{t-1} = x_{t-2} + u_t$ is the lagged original time series again of the same order of interation. In fact $x_t \in I(\infty)$.

[5 marks]

Question 4.

4. A student having data for time series (1979-2014) on expenditures of US citizens on doctors *DOC*, treatment in hospitals *HOSP*, dentists *DENT*, ophthalmologists *OPHT*, drug preparations *PHAR*, health, and life insurance: *HINS* and *LIFE*, and recreation *REC*, considers them together as a panel data (as observations for particular units of some general type of good – medical services *MED*). The student is interested in evaluation of income elasticity and relative price elasticity on the base of the panel data model. She begins from fixed effects model (LSDV method) and obtains the following results:

$$\ln \hat{MED}_{it} = -6.17 + 1.33 \cdot \ln DPI_t - 0.25 \cdot \ln PRMED_{it} \quad R^2 = 0.99$$
$$(0.22)\ (0.025) \qquad (0.056)$$

with the following fixed effects $DOC(1.15)$, $HOSP(1.54)$, $DENT(0.02)$, $OPHT(-1.56)$, $HINS(-0.23)$, $LIFE(0.16)$, $PHAR(-1.58)$, $REC(0.50)$.

(a) Describe briefly the LSDV method and explain why it could help to solve the problem of unobserved heterogeneity for panel data model. Interpret the coefficients of the model and explain the meaning of fixed effects.

Solution

The coefficients of variables are correspondingly income and relative price elasticities of expenditures on medical services. It could be noted that income elasticity is high (the value of 1.33 is greater than unity) while price elasticity is low in absolute value

The main question in the panel data analysis is the problem of the origin of unobserved heterogeneity $\sum_{p=1}^{s} \gamma_p Z_{pi}$ = α_i in the model of the type

$$Y_{it} = \beta_0 + \sum_{j=1}^{k} \beta_j X_{jit} + \sum_{p=1}^{s} \gamma_p Z_{pi} + \varepsilon_{it} \qquad (1).$$

Fixed effect approach assigns it to the fixed characteristics of individual elements, while random effect approach assigns it to the random factors.

In **Least Squares Dummy Variable (LSDV) fixed effect method** the unobserved effect is brought explicitly into the model. A set of dummy variables D_i is defined, where D_i is equal to 1 in the case of an observation relating to an individual *i* and 0 otherwise. The model can be written as

$$Y_{it} = \sum_{j=2}^{K} \beta_j X_{jit} + \sum_{i=1}^{n} \alpha_i D_i + u_{it}. \qquad \text{(vi)}$$

The unobserved effect is now being treated as the coefficient of the specific individual *i*.

The term $\alpha_i D_i$ represents a fixed effect on the dependent variable Y_i for individual *i*.

If we want to keep the intercept in the model then instead of *n* dummy variables $(n-1)$ dummy variables have to be used, otherwise we will fall into the dummy variable trap. Inclusion of a set of dummies leads to losing of additional degrees of freedom, so estimators becomes less efficient.

(b) What are other methods of fixed effects data panel model? What are their comparative advantages and disadvantages?

Solution.
In **Within-groups fixed effect** method for the initial model

$$Y_{it} = \beta_1 + \sum_{j=2}^{K} \beta_j X_{jit} + \alpha_i + u_{it} \,;\; i = 1, 2, ..., n \,;\; t = 1, 2, ..., T \,. \qquad \text{(i)}$$

we first evaluate the averages summing up the observations of each cross-sectional unit over the time dimension and divide by T to get

$$\overline{Y}_i = \beta_1 + \sum_{j=2}^{K} \beta_j \overline{X}_{ij} + \alpha_i + \overline{u}_i \,. \qquad \text{(ii)}$$

and then subtracting (ii) from (i), we get model in deviations from the means

$$Y_{it} - \overline{Y}_i = \sum_{j=2}^{K} \beta_j (X_{jit} - \overline{X}_{ij}) + u_{it} - \overline{u}_i \,. \qquad \text{(iii)}$$

In (iii) we can see that the unobserved effect (α_i) disappears. This is known as within-groups regression. It explains the variations about the mean of the dependent variable in terms of the variations about the means of the explanatory variables for the group of observations relating to a given individual.
Drawbacks: it can be shown that the within-groups method is identical to the LSDV method. Hence, the drawbacks are the same.

• The intercept and any explanatory variable that remain constant for each individual will drop out of the model.

• The variation in $(X_{ij} - \overline{X}_i)$ may be much smaller than the variation in X_j. If this is the case, the impact of the disturbance term may be relatively large, giving rise to imprecise estimates.

• There is loss of a substantial number of degrees of freedom.

In **First Differences fixed effect** approach, the unobserved heterogeneity is eliminated by subtracting the observation from the previous time period from the observation for the current time period, for all time periods.
Lag (i) by one period to get

$$Y_{it-1} = \beta_1 + \sum_{j=2}^{K} \beta_j X_{jit-1} + \alpha_i + u_{it-1} \,. \qquad \text{(iv)}$$

Subtracting (i) from (iv), we get

$$\Delta Y_{it-1} = \sum_{j=2}^{K} \beta_j \Delta X_{jit-1} + \alpha_i + u_{it} - u_{it-1} \qquad \text{(v)}$$

and the unobserved heterogeneity (α_i) disappears.
Drawbacks:
1) The intercept and any explanatory variable that remain constant for each individual will drop out of the model.
2) n degrees of freedom are lost as the first observation for each individual is not defined.
3) It gives rise to additional autocorrelation of the moving average type.

(c) The student runs also random effect model and then perform Durbin-Wu-Hausman test to choose the best model. The results of the test are following

Hausman Test

Test Summary	Statistic	d.f.	Prob.
Cross-section random	0.8564	2	0.6517

What are assumptions of the Hausman test? What are possible conclusions and their advantages? What could be final conclusion here?

Solution.

For application of the random effect approach one should be possible to treat each of the unobserved Z_p variables as being drawn randomly from the same distribution. Moreover it is supposed that the Z_p variables are distributed independently of all of the X_j variables.

To decide which approach is more appropriate one should use Durbin–Wu–Hausman (DWH) test, that uses a statistic based on chi-square distribution with the number of degrees of freedom equal to the number of explanatory variables X_j in the equation (1) (equal to 2 in our case: $\log(DPI_t)$ and $\log(PRMED)$). The DWH test determines whether the estimates of the coefficients, taken as a group, are significantly different in the two regressions. The null hypothesis: the α_i are distributed independently of the X_j. In our case the null hypothesis is not rejected at 5% level (as p-value=0.6517 > 0.05). On the other hand an additional argument in favor of fixed effect model is that several branches of medicine represented in the sample hardly be considered as random representative of medicine in general.

Advantages and risks of the possible conclusions

If the null hypothesis is correct, both random effects and fixed effects are consistent, but fixed effects will be inefficient because, it involves estimating an unnecessary set of dummy variable coefficients.

If the null hypothesis is correct, random effects is more attractive because observed characteristics that remain constant for each individual are retained in the regression model. With random effects estimation we do not lose *n* degrees of freedom, as is the case with fixed effects.

If the null hypothesis is false, the random effects estimates will be subject to unobserved heterogeneity bias and will therefore differ systematically from the fixed effects estimates.

Additional material (gives bonus marks)

If fixed effect approach or random effect approach is chosen there is always a possibility that there is no unobserved heterogeneity at all and one can apply pooled regression approach. To discriminate these cases there are some tests for this purpose, for example Breush-Pagan test based on Lagrange Multiplier approach. It uses also some chi-square distribution with $df = 1$ under H_0 of the absence of random effects.

[5 marks]

SECTION B
(1 hour 10 minutes)
Answer **TWO** questions from this section (questions **5-7**).
Each answered question in this section bears 20 marks.

Question 5.

5. Almost 105 years ago on April 14, 1912, the unthinkable happened when the "unsinkable" Royal Mail Ship (RMS) Titanic crashed into an iceberg and sunk into the Atlantic Ocean. The 20 lifeboats aboard the ship were not enough to save a majority of the passengers, leaving over 1500 passengers and crew members aboard the sinking vessel. A total of 705 passengers escaped onto lifeboats and to safety. But not all passengers had an equal chance of getting onto a lifeboat and surviving the disaster.
A student has found in internet the data on 2201 passengers and crew members of RMS Titanic during his last voyage across the ocean. She is trying now, using regression analysis, to estimate the impact of passenger class (***CLASS***=0 for the crew members, ***CLASS***=1 for the first class passengers, ***CLASS***=2 for the second class passengers, ***CLASS***=3 for the third class passengers), gender (***MALE***=1 for male and 0 for female) and age (***AGE***=1 for adults and 0 for children) on a person's likelihood of surviving the shipwreck ***SURV***. In the data ***SURV***=1 if the person survived the shipwreck and ***SURV***=0 if not. She runs different regressions with the following results (dependent variable ***SURV*,** regression (i) uses full sample, other regressions use subsample without crew members; (asymptotical) standard errors in parentheses)**.**

	(i) (OLS)	**(ii)** (OLS)	**(iii)** (LOGIT)	**(iv)** (LOGIT)	**(v)** (PROBIT)
CLASS	−0.0515 (0.0072)	−0.1515 (0.0134)	−0.489 (0.043)	−0.875 (0.085)	−0.502 (0.048)
AGE	−0.165 (0.007)	−0.181 (0.040)	—	−1.056 (0.243)	−0.580 (0.138)
MALE	−0.552 (0.022)	−0.478 (0.023)	—	−2.367 (0.145)	−1.415 (0.084)
Constant	0.985 (0.047)	1.208 (0.053)	0.793 (0.103)	3.895 (0.347)	2.257 (0.193)
N	2,201	1,316	1,316	1,316	1,316
R^2	0.228	0.331			
McFadden R^2			0.0758	0.269	0.267
RSS	371.59	207.31	278.52	204.30	205.21
LogL	−1165.43	−651.45	−807.15	−638.40	−639.93
LR stat,			132.45	469.95	466.90

(a) Give interpretation to the regressions **(i)** and **(ii)**. Are they significant? What are the advantages and the disadvantages of these models? Why in models **(iii)-(iv)-(v)** these shortcomings are absent or mitigated?

Solution
Regressions **(i)** and **(ii)** represent linear probability model (LPM), where ***SURV*** shows the probability to survive. The coefficients of these models could be interpreted directly as marginal effects. For example, the coefficient of ***CLASS*** -0.0515 shows that second-class passengers (other factors being equal) have a chance of survival about 5 percentage points less than first-class passengers (and so the third-class passengers as to compare with the second-class passengers. For children, the probability of survival is higher by about 16

percentage points. The model shows that when other factors unchanged, the probability of survival for women is about 50 percentage points higher than for a man.
All coefficients are significant, while R-squared is not high indicating that there are some other factors not included into equation.
The advantage of LPM is its simplicity that allows to show general picture, while many its drawbacks prevents to believe that OLS estimators are correct and tests are valid.
First linear probability model can produce predictions outside the range $[0,1]$ which is not feasible. We can observe this in regression (ii) where probability to survive for female not adult person of any class exceeds unity.
Also the error terms are heteroskedastic since

$$u_i = 1 - \beta_1 - \beta_2 CLASS - \beta_3 AGE - \beta_4 MALE \text{ with probability } p$$

and

$$u_i = -\beta_1 - \beta_2 CLASS - \beta_3 AGE - \beta_4 MALE \text{ with probability } (1-p).$$

The distribution of the error term is certainly not normal so usual tests strictly speaking could not be applied.
Additional drawback of the model is the constant marginal effect of the explanatory variable.
Some of these drawbacks can be overcome by using binary choice models. For example the logit model refers to a model where the probability of the occurrence of the event is determined by the logistic function, $F(z) = 1/(1+e^{-z})$, where $z_i = \beta'_1 + \beta'_2\, CLASS_i + \beta'_3\, AGE_i + \beta'_4\, MALE_i + u'_i$. The function is confined within the range $[0,1]$ as can be seen by allowing z to tend to $+\infty$ or $-\infty$ and therefore does not suffer from the same problems as the linear probability model. It can be relatively easily estimated by maximum likelihood techniques therefore the estimates have the standard maximum likelihood properties i.e. the estimators are consistent, asymptotically efficient and asymptotically normally distributed.
[5 marks]

(b) Using regression **(iv)** help the student to evaluate the marginal effect of passenger class, taking as the initial values an adult female person traveling in a first class cabin (use method of derivatives). Compare your result with the corresponding result for regression **(ii)**

Solution.
Regressions **(iv)** is LOGIT regression of the type $F(z) = 1/(1+e^{-z})$,
where $z_i = \beta'_1 + \beta'_2\, CLASS_i + \beta'_3\, AGE_i + \beta'_4\, MALE_i + u'_i$, in our case
$SURV_i = p_i = F(Z_i) = \dfrac{1}{1+e^{-Z_i}}$, *Z*=2.69 - 0.875**CLASS* - 1.056**AGE* - 2.367**MALE*. For the person under consideration (an adult female person traveling in a second class cabin)
Z=3.895 - 0.875*1 - 1.056*1 - 2.367*0 = 1.964. To evaluate marginal effect we need to use chain rule for evaluation derivative of $SURV_i = p_i$ with respect to $X_1 = CLASS$: $\dfrac{\partial p}{\partial X_i} = \dfrac{dp}{dZ}.\dfrac{\partial Z}{\partial X_1} = f(Z)\cdot\beta_1$. Now $f(Z) = \dfrac{dp}{dZ} = \dfrac{e^{-Z}}{(1+e^{-Z})^2}$ that is $\dfrac{e^{1.964}}{(1+e^{1.964})^2} = 0.108$ and $\beta_1 = -0.875$, so $\dfrac{\partial p}{\partial X_i} = f(Z)\cdot\beta_1 = 0.2401*(-0.875) = -0.0945$. This value is comparable with -0.1515 - marginal effect of *CLASS* in Linear Probability Model **(ii)** for the same sample. The difference is explained by the fact that in the LPM the limiting effect is constant, whereas in the LOGIC model it is specific for each individual.
[5 marks]

(c) What are McFadden R^2 and *LR* statistic? How they can be used for evaluation of the statistical quality of the regression? Comparing regressions **(iii)** and **(iv)** evaluate whether two dummy variables ***AGE*** and ***GENDER*** are significant both separately and taken together?

Solution

McFadden R-squared is evaluated using formula $pseudo-R^2 = 1 - \frac{\log L}{\log L_0}$ where $\log L$ is the value of log likelihood for the model and $\log L_0$ is the same for the model which included only constant. This coefficient is like R-squared alwais between 0 and 1, but unlike the conventional R-squared it could not be interpreted as the percentage of explained variance of the dependent variable.

Log likelihood is the value of a function of log of the probability function used in the method of MLE. As likelihood means probability its logarithm is always negative.

The LR statistics is so called likelyhood ratio $LR = 2\log\frac{L}{L_0} = 2(\log L - \log L_0)$ could be used for testing general explanatory power of the model. It is distributed os a chi-squared statistics with $k-1$ degrees of freedom, where $k-1$ is the number of explanatory variables, under the null hypotheses that the coefficients of the variables are jointly equal to 0

To estimate significance of each variable separately can be done using tables of z-distribution; z-statistics here are in fact some asymptotic values that allows for large samples to use the tables of normal distribution to evaluate the significance of the coefficients. For example in regression **(iv)** variable ***AGE*** is significant as $z = \left|\frac{-0.580}{0.138}\right| = 4.2 > 2{,}57 = z_{crit}(1\%)$.

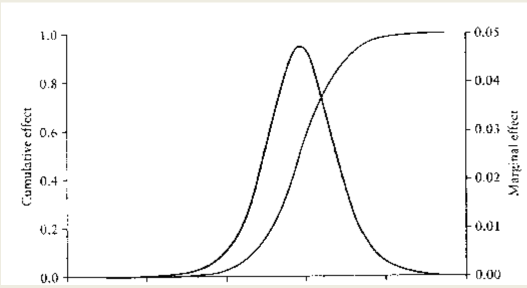
The estimation of the joint significance of all variables could be performed using LR statistic that has chi-square distribution with 2 degrees of freedom (number of restrictions). From regression output for **(iii)** and **(iv)** $LR = 2(\log L_U - \log L_R) = 2(-638.4 + 807.15) = 337.5 > 9.210 = \chi^2_{crit}(1\%)$.

[5 marks]

(d) What is the difference between regressions **(iv)** and **(v)?** Help the student to understand whether for female passengers of Titanic the chance to survive is significantly greater. To do this for regression **(v)** evaluate the marginal effect of gender, taking as the initial values an adult male person traveling in a first class cabin (use direct calculation without derivatives). Compare your result with the corresponding result for regression **(ii)**.

Solution

The Probit model is based on the distribution function of normal distribution $F(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{x^2}{2}}dx$ that also form a sigmoid curve making marginal effect being not constant



The results obtain using LOGIT and PROBIT models usually quite similar. PROBIT model is convenient for direct evaluation of marginal effects as the value of $F(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{x^2}{2}}dx$ is tabulated (the table of cumulative normal distribution). So using PROBIT regression

$F(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{x^2}{2}}dx$, where $z_i = \beta'_1 + \beta'_2 CLASS_i + \beta'_3 AGE_i + \beta'_4 MALE_i + u'_i$.

evaluate value of z using regression **(v)** $z_i = 2.257 - 0.502 CLASS_i - 0.580 AGE_i - 1.415 MALE_i + u'_i$.

For the person under consideration (an adult male person traveling in a second class cabin):

$z_i = 2.257 - 0.502 \cdot 1 - 0.580 \cdot 1 - 1.415 \cdot 1 = -0.24$. From the tables of normal distribution $F(-0.24) = 1 - 0.591 = 0.409$. To evaluate marginal effect of gender let us change ***MALE*** to 0:
$z_i = 2.257 - 0.502 \cdot 1 - 0.580 \cdot 1 - 1.415 \cdot 0 = 1.175$. Now $F(1.175) = 0.9599$, so the marginal effect of gender (additional chance to survive) is $F(1.175) - F(-0.24) = 0.9599 - 0.409 = 0.5509$. Again it is close to LPM estimate 0.478. So really the key factor of survival was the gender of the passenger (at least for that time!): ladies had more than 50 additional percentage points for chance to survive than gentlemen.
[5 marks]

Question 6.

6. Consider an $ADL(2,1)$ model:

$$Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 Y_{t-2} + \alpha_4 X_t + \alpha_5 X_{t-1} + u_t;\ t = 1, 2, ..., T. \quad (1)$$

where both Y_t and X_t are $I(1)$, u_t is the disturbance term where $\mathrm{E}(u_t) = 0$; $\mathrm{E}(u_t^2) = \sigma^2$ and $\mathrm{E}(u_s u_t) = 0$ for any $s \neq t$. We will assume that all variables are in logarithmic form so, for example α_4 can be interpreted as a short term *X*-elasticity of *Y*.

(a) Explore long run dynamics of this model and find the long term *X*-elasticity of *Y*. How would you estimate a cointegrating relationship between Y_t and X_t?

Solution

Suppose that variables Y_t and X_t both tends to their long run steady states $\bar{Y}$ and $\bar{X}$. Then

$Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 Y_{t-2} + \alpha_4 X_t + \alpha_5 X_{t-1} + u_t$ (1)

converts to

$\bar{Y} = \alpha_1 + \alpha_2 \bar{Y} + \alpha_3 \bar{Y} + \alpha_4 \bar{X} + \alpha_5 \bar{X}$.

From here

$\bar{Y} = \dfrac{\alpha_1}{1 - \alpha_2 - \alpha_3} + \dfrac{\alpha_4 + \alpha_5}{1 - \alpha_2 - \alpha_3} \bar{X}$. (2)

As both $\bar{Y}$ and $\bar{X}$ are in logarithmic form the expression $\dfrac{d\bar{Y}}{d\bar{X}} = \dfrac{\alpha_4 + \alpha_5}{1 - \alpha_2 - \alpha_3}$ represent the long term *X*-elasticity of *Y*.
[5 marks]

(b) Transform the model (1) into the one relating to the first differences of *X* and *Y*. What are advantages and disadvantages of this model?

Solution

Lagging equation

$Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 Y_{t-2} + \alpha_4 X_t + \alpha_5 X_{t-1} + u_t$ (1)

we get

$Y_{t-1} = \alpha_1 + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3} + \alpha_4 X_{t-1} + \alpha_5 X_{t-2} + u_t$

and subtracting it from original equation (1) we obtain the model in difference form

$\Delta Y_t = \alpha_2 \Delta Y_{t-1} + \alpha_3 \Delta Y_{t-2} + \alpha_4 \Delta X_t + \alpha_5 \Delta X_{t-1} + u_t - u_{t-1}$ (3)

As both Y_t and X_t are $I(1)$ all variables of the model (3) are stationary, the regression model can be used for analysis , while regression (1) based on the non stationary time series could be spurious.
Om the other hand the model (3) based only on differences could be shortsighted and so not sutable for analysis and forecasting. Moreover its disturbance term $\varepsilon_t = u_t - u_{t-1}$ is of the moving average type and so could suffer

from (negative) autocorrelation. It could generally detoriate the estimators of the model's coefficients making them less efficient.
Note that if original model (1) suffers from positive autocorrelation the imposing of positive and negative autocorrelations can help to get rid of autocorrelation.
[5 marks]

(c) Supposing that the cointegrating relationship between *X* and *Y* had been found, express the model (1) as an error correction type model, and describe briefly its dynamics.

Solution
First subtract Y_{t-1} from both sides of equation (1)
$$Y_t - Y_{t-1} = \alpha_1 + \alpha_2 Y_{t-1} - Y_{t-1} + \alpha_3 Y_{t-2} + \alpha_4 X_t + \alpha_5 X_{t-1} + u_t:$$
Then simultaneously add to and subtract from right side $\alpha_3 Y_{t-1}$
$$Y_t - Y_{t-1} = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 Y_{t-1} - Y_{t-1} - \alpha_3 Y_{t-1} + \alpha_3 Y_{t-2} + \alpha_4 X_t + \alpha_5 X_{t-1} + u_t:$$
or
$$\Delta Y_t = \alpha_1 + (\alpha_2 + \alpha_3 - 1) Y_{t-1} - \alpha_3 \Delta Y_{t-1} + \alpha_4 X_t + \alpha_5 X_{t-1} + u_t:$$
Do the same with $\alpha_4 X_{t-1}$
$$\Delta Y_t = \alpha_1 + (\alpha_2 + \alpha_3 - 1) Y_{t-1} - \alpha_3 \Delta Y_{t-1} + \alpha_4 X_t + \alpha_4 X_{t-1} - \alpha_4 X_{t-1} + \alpha_5 X_{t-1} + u_t:$$
Rearrange terms
$$\Delta Y_t = (\alpha_2 + \alpha_3 - 1) Y_{t-1} + \alpha_1 + (\alpha_4 + \alpha_5) X_{t-1} - \alpha_3 \Delta Y_{t-1} - \alpha_4 \Delta X_t + u_t:$$
or
$$\Delta Y_t = (\alpha_3 + \alpha_2 - 1)(Y_{t-1} - \frac{\alpha_1}{1-\alpha_2-\alpha_3} - \frac{\alpha_4+\alpha_5}{1-\alpha_2-\alpha_3} X_{t-1}) - \alpha_3 \Delta Y_{t-1} + \alpha_4 \Delta X_t + u_t$$
Error correction term $Y_{t-1} - \frac{\alpha_1}{1-\alpha_2-\alpha_3} - \frac{\alpha_4+\alpha_5}{1-\alpha_2-\alpha_3} X_{t-1}$ coinsides in its form with cointegrating relationship (2) and its coefficient $\frac{\alpha_4+\alpha_5}{1-\alpha_2-\alpha_3}$ is long term *X*-elasticity of , the coefficient $\alpha_3 + \alpha_2 - 1$ shows the speed of correction.

Alternative solution (Johnston, DiNardo, Econometric Methods, 4th ed., page 247)
Starting from
$$Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 Y_{t-2} + \alpha_4 X_t + \alpha_5 X_{t-1} + u_t \quad (1)$$
Let us perform reparametrization: $Y_t = Y_{t-1} + (Y_t - Y_{t-1})$, $Y_{t-2} = Y_{t-1} - (Y_{t-1} - Y_{t-2})$, $X_t = X_{t-1} + (X_t - X_{t-1})$, or $Y_t = Y_{t-1} + \Delta Y_t$, $Y_{t-2} = Y_{t-1} + \Delta Y_{t-1}$, $X_t = X_{t-1} + \Delta X_t$ and substitute these expression into model (1)
$$Y_{t-1} + \Delta Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 (Y_{t-1} - \Delta Y_{t-1}) + \alpha_4 (X_{t-1} + \Delta X_t) + \alpha_5 X_{t-1} + u_t \quad \text{or}$$
$$\Delta Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 Y_{t-1} - Y_{t-1} - \alpha_3 \Delta Y_{t-1} + \alpha_4 X_{t-1} + \alpha_4 \Delta X_t + \alpha_5 X_{t-1} + u_t \quad \text{so}$$
$$\Delta Y_t = (\alpha_2 + \alpha_3 - 1) Y_{t-1} + \alpha_1 + (\alpha_4 + \alpha_5) X_{t-1} - \alpha_3 \Delta Y_{t-1} + \alpha_4 \Delta X_t + u_t \quad \text{so}$$
$$\Delta Y_t = (\alpha_2 + \alpha_3 - 1)(Y_{t-1} - \frac{\alpha_1}{1-\alpha_2-\alpha_3} - \frac{(\alpha_4+\alpha_5)}{1-\alpha_2-\alpha_3} X_{t-1}) - \alpha_3 \Delta Y_{t-1} + \alpha_4 \Delta X_t + u_t.$$

[5 marks]

(d) Express the model (1) as the error correction type model with two equations and unobserved variable *Y**, interpret its coefficients.

Solution
We start now from error correction model in **(c)**:

$$\Delta Y_t = (\alpha_3 + \alpha_2 - 1)(Y_{t-1} - \frac{\alpha_1}{1-\alpha_2-\alpha_3} - \frac{\alpha_4+\alpha_5}{1-\alpha_2-\alpha_3} X_{t-1}) - \alpha_3 \Delta Y_{t-1} + \alpha_4 \Delta X_t + u_t$$

Let us denote $\lambda = 1 - \alpha_3 - \alpha_2$, $\beta_1 = \frac{\alpha_1}{1-\alpha_2-\alpha_3}$ and $\beta_2 = \frac{\alpha_4+\alpha_5}{1-\alpha_2-\alpha_3}$, then

$$\Delta Y_t = -\lambda(Y_{t-1} - \beta_1 - \beta_2 X_{t-1}) - \alpha_3 \Delta Y_{t-1} + \alpha_4 \Delta X_t + u_t$$

As $Y_t^* = \beta_1 + \beta_2 X_t$ then

$\Delta Y_t = -\lambda(Y_{t-1} - Y_t^*) - \alpha_3 \Delta Y_{t-1} + \alpha_4 \Delta X_t + u_t$ or finally

$$\Delta Y_t = \lambda(Y_t^* - Y_{t-1}) - \alpha_3 \Delta Y_{t-1} + \alpha_4 \Delta X_t + u_t$$

So the model

$$Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 Y_{t-2} + \alpha_4 X_t + \alpha_5 X_{t-1} + u_t$$

is can be represented as an error correction model un which error correction term $Y_t^* - Y_{t-1}$ is powered by the target equation $Y_t^* = \beta_1 + \beta_2 X_t$ and the Y_t follows the partial adjustment process $Y_t = \lambda Y_t^* + (1-\lambda) Y_{t-1}$.

Coefficient $\beta_2 = \frac{\alpha_4+\alpha_5}{1-\alpha_2-\alpha_3}$ is the long term *X*-elasticity of *Y*, $\lambda = 1 - \alpha_2 - \alpha_3$ is a speed of adjustment in the process $Y_t = \lambda Y_t^* + (1-\lambda) Y_{t-1}$, and also the strenght of the error correction in the model

Alternative solution

According to the assumptions of the partial adjustment model the dynamics of a variable follows Y_t the adjustment process of

$$Y_t = \lambda Y_t^* + (1-\lambda) Y_{t-1}$$

where $Y_t^* = \beta_1 + \beta_2 X_t$ is a target or desired state for the moment t (unobservable variable). The initial process

$Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 Y_{t-2} + \alpha_4 X_t + \alpha_5 X_{t-1} + u_t$ (1)

could be represented in a form of combination of difference model and partial ajustment model of the type:

$$\Delta Y_t = \lambda(Y_{t-1}^* - Y_{t-1}) + \delta \Delta Y_{t-1} + \gamma \Delta X_t + u_t$$

The simplest way to do this is to open brakets in

$$Y_t - Y_{t-1} = \lambda(Y_{t-1}^* - Y_{t-1}) + \delta(Y_{t-1} - Y_{t-2}) + \gamma(X_t - X_{t-1}) + u_t:$$

$$Y_t = \lambda Y_{t-1}^* - \lambda Y_{t-1} + Y_{t-1} + \delta Y_{t-1} - \delta Y_{t-2} + \gamma X_t - \gamma X_{t-1} + u_t.$$

and substitute expression for $Y_{t-1}{}^*$ from lagged target equation $Y_{t-1}^* = \beta_1 + \beta_2 X_{t-1}$:

$$Y_t = \lambda\beta_1 + \lambda\beta_2 X_{t-1} - \lambda Y_{t-1} + Y_{t-1} + \delta Y_{t-1} - \delta Y_{t-2} + \gamma X_t - \gamma X_{t-1} + u_t.$$

Rearranging terms

$$Y_t = \lambda\beta_1 + (\delta - \lambda - 1) Y_{t-1} - \delta Y_{t-2} + \gamma X_t + (\lambda\beta_2 - \gamma) X_{t-1} + u_t$$

and comparing them with the initial equation (1)

$Y_t = \alpha_1 + \alpha_2 Y_{t-1} + \alpha_3 Y_{t-2} + \alpha_4 X_t + \alpha_5 X_{t-1} + u_t$ (1)

we get a system of equations

$$\lambda\beta_1 = \alpha_1,\ \delta - \lambda + 1 = \alpha_2,\ -\delta = \alpha_3,\ \gamma = \alpha_4,\ \lambda\beta_2 - \gamma = \alpha_5$$

from which we easily obtain a solution

$$\gamma = -\alpha_4;\ \delta = -\alpha_3;\ \lambda =_2 1 - \alpha_2 + \delta = 1 - \alpha_2 - \alpha_3;\ \beta_1 = \frac{\alpha_1}{\lambda} = \frac{\alpha_1}{1-\alpha_2-\alpha_3};\ \ \beta_2 = \frac{\gamma+\alpha_5}{\lambda} = \frac{\alpha_4+\alpha_5}{1-\alpha_2-\alpha_3}$$

Here again $\beta_2 = \frac{\alpha_4+\alpha_5}{1-\alpha_2-\alpha_3}$ is the long term *X*-elasticity of *Y*, $\lambda = 1 - \alpha_2 - \alpha_3$ is a speed of adjustment in the process $Y_t = \lambda Y_t^* + (1-\lambda) Y_{t-1}$, and also the strenght of the error correction in the model

$$Y_t - Y_{t-1} = \lambda(Y_{t-1}^* - Y_{t-1}) + \delta(Y_{t-1} - Y_{t-2}) + \gamma(X_{t-1} - X_{t-2}) + u_t:$$

[5 marks]

Partial solution

We start again from error correction model in **(c)**:

$$\Delta Y_t = (\alpha_3 + \alpha_2 - 1)(Y_{t-1} - \frac{\alpha_1}{1-\alpha_2-\alpha_3} - \frac{\alpha_4+\alpha_5}{1-\alpha_2-\alpha_3} X_{t-1}) - \alpha_3 \Delta Y_{t-1} + \alpha_4 \Delta X_t + u_t$$

and substitute the value of Y^*_{t-1} instead of Y_{t-1} from lagged partial adjustment equation $Y_t = \lambda Y^*_t + (1-\lambda)Y_{t-1}$ where $Y^*_t = \beta_1 + \beta_2 X_t$.

This approach describes only the general idea of solution, we need also to show that in fact

$\lambda = 1 - \alpha_3 - \alpha_2$, $\beta_1 = \dfrac{\alpha_1}{1-\alpha_2-\alpha_3}$ and $\beta_2 = \dfrac{\alpha_4+\alpha_5}{1-\alpha_2-\alpha_3}$

[2 marks]

Question 7.

A student decided to investigate the behaviour of the exchange rates of Russian ruble to EURO and to GPB - Great Britain Pound (Rubles per Euro RUR_EUR, and Rubles per GBP RUR_GBP) in the period 31.03.2015 – 28.03.2017. She had downloaded 494 daily observations from the Central Bank of Russia website (correctly deciding to neglect some irregularities due to missing data for national holidays).

First, the student tested the series and their first differences D(…) for stationarity, doing Augmented Dickey-Fuller tests:

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RUR_GBP)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RUR_GBP(-1)	-0.003542	0.004777	-0.741427	0.4588
D(RUR_GBP(-1))	0.028780	0.045174	0.637087	0.5244
C	0.290414	0.432204	0.671936	0.5019

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RUR_GBP,2)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(RUR_GBP(-1))	-1.037785	0.062814	-16.52161	0.0000
D(RUR_GBP(-1),2)	0.065743	0.044969	1.461968	0.1444
C	-0.029105	0.055758	-0.521984	0.6019

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RUR_EUR)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RUR_EUR(-1)	-0.009126	0.006134	-1.487775	0.1375
C	0.638373	0.432333	1.476579	0.1404

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RUR_EUR,2)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(RUR_EUR(-1))	-1.028017	0.045056	-22.81651	0.0000
C	-0.001235	0.045875	-0.026921	0.9785

(a) Explain the specification differences in the equations estimated and formulate the H_0 and H_1 hypotheses.

Solution

a) The second specification corresponds to usual autoregression process $Y_t = \beta_1 + \beta_2 Y_{t-1} + \varepsilon_t$ and uses the equation $Y_t - Y_{t-1} = (\beta_2 - 1)Y_{t-1} + \varepsilon_t$. Null hypothesis is $H_0 : \beta_2 = 1$ (random walk – non stationary time series) againts $H_0 : -1 < \beta_2 < 1$ (stationary autoregressive process). The first case corresponds to the more complex model

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 Y_{t-2} + \varepsilon_t .$$

Subtract Y_{t-1} from the both sides, add and subtract $\beta_3 Y_{t-1}$ on the right side

$$Y_t - Y_{t-1} = \beta_1 + \beta_2 Y_{t-1} + \beta_3 Y_{t-1} - Y_{t-1} + \beta_3 Y_{t-2} - \beta_3 Y_{t-1} + \varepsilon_t .$$

We get

$$\Delta Y_t = \beta_1 + (\beta_2 + \beta_3 - 1)Y_{t-1} - \beta_3 \Delta Y_{t-1} + \varepsilon_t .$$

For this process the Augmented Dickey-Fuller test is applicable

Null hypothesis is $H_0 : \beta_2 + \beta_3 = 1$ (non-stationary time series) againts $H_0 : -1 < \beta_2 + \beta_3 < 1$ (stationary generalized autoregressive process). The optimum lag length could be chosen on the base of some information criteria (Akaike or Schwarz).

[5 marks].

(b) Then the student decided to test if the series RUR_GBP and RUR_EUR are cointegrated, and discovered that they are not.

Comment on the stationarity of each of two series and their first differences, and find their orders of integration. Describe the concept of cointegration, and the procedure she did.

Solution

Two time series are cointegrated if

1) they are of the same order of integration greater or equal than 1, say I(1);

2) there is a linear combination of them that is stationary.

The condition 1 is implemented here. The second condition is equivalent to the following: the residuals of the regression of one of the time series on the other are stationary. Then the student should regress RUR_GBP on RUR_EUR (or vice versa) and investigate whether the residuals of this regression are stationary (standard procedure to be discussed, with the indication that the special critical values are taken because we teas the residuals instead of the disturbance term).

The RUR_GBP process is non-stationary as null hypothesis of random walk is not rejected at 5% level (-0.741 > -2.87). The RUR_EUR is also non-stationary at 5% level since
-1.4878>-2.87.

The first differences of both time series are stationary since -16.52 and -22.82 are both less than -3.44, and the null hypothesis of non-stationarity is rejected even at the 1% level.

Comment: one should use critical values and not rely to p-value in the table as distribution of test statistic here is different from conventional t-statistics.The df is different in the model 2, but since it is large, the critical values are the same with the rounding used.

[5 marks].

(c) Then the student decided to estimate the AR(1) and ADL(1,1) regressions and do the Common Factor test. For these two regressions the Residual Sums of Squares are 109.806 and 109.542 respectively. Calculate the test statistic (hint: $\log(1+\alpha) \approx \alpha$ for small α) and do the test. Which of two specifications would you choose, and why?

Solution

Let X_t and Y_t be two time series under consideration. If we assume that $Y_t = \beta_1 + \beta_2 X_t + u_t$ is AR(1) regression, then $u_t = \rho u_{t-1} + \varepsilon_t$. Using autoregressive transformation the autocorrelation could be removed from data.

It is also possible to transform the AR(1) model into specification $Y_t = \beta_1(1-\rho) + \beta_2 X_t - \beta_2 \rho X_{t-1} + \rho Y_{t-1} + \varepsilon_t$ (the transformation to be described).

This is a special form for the model ADL(1,1) $Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 Y_{t-1} + \varepsilon_t$ under nonlinear restriction $\beta_3 = -\beta_2 \beta_4$. Testing this restriction (so called Common Factor test) uses chi-square statistic $n \ln \frac{RSS_1}{RSS_0} = (494-1) \cdot \ln \frac{109.806}{109.542} = 493 \cdot \ln(1 + 0.00241) \approx 493 \cdot 0.00241 = 1.188$ what is less than critical value of 3.8415 (for the degrees of freedom equal to the number of restrictions =1). So the restriction is not rejected. As this restriction could be derived from assumption of autocorrelated disturbance term $u_t = \rho u_{t-1} + \varepsilon_t$ it means that ρ (coefficient of y_{t-1} could be interpreted as autocorrelation coefficient. The AR(1) model would be preferred to the ADL(1,1) model since the estimates in the model with valid restriction are more efficient.

[5 marks].

(d) After that, the student reminded that she forgot to test the residuals for autocorrelation, and decided to test the ADL(1,1) model first. Explain why and how she could apply the Breusch-Godfrey LM test, with lag=2, for it.

Solution

The Breusch–Godfrey test allows to detect the residuals autocorrelation of the order higher than 1. It is a large samples test. It is applicable if the lagged dependent variable is one of the regressors which is important to the model considered. The most common realization of Breusch–Godfrey test involves the computation of the Lagrange multiplier statistic nR^2 when the residuals regression is fitted, n being the actual number of observations in the regression.

Let regression equation be $Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 Y_{t-1} + \varepsilon_t$. For the case of lag=2, we run this regression we find residuals e_t, and using them run auxiliary regression

$$\hat{e}_t == \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 Y_{t-1} + \rho_1 e_{t-1} + \rho_2 e_{t-2}.$$

Test statistic: nR^2, distributed as $\chi^2(2)$ when testing for second order autocorrelation. Alternatively, simple F test on coefficients of e_{t-1} and e_{t-2} , again with asymptotic validity. Here df=1034-1-2=1031 since 1 observation is lost in the ADL(1,1) model, and 2 in the auxiliary regression with lag=2.

[5 marks]

The International College of Economics and Finance
Econometrics – 2017-2018.
Midterm exam 2018. March 29.

<u>General instructions.</u> Candidates should answer SIX of the following SEVEN questions: all 4 questions of the Section A and any 2 of the questions from Section B (questions 5-7). The weight of the Section A is 60% of the exam; three other questions from the Section B add 20% each. You are advised to divide your time accordingly. Structure your answers in accordance with the structure of the questions. When testing hypotheses always state clearly null and alternative hypotheses provide critical value used for the test, mentioning degrees of freedom and the significance level chosen for the test.

Warning: the proposed solutions are not the only correct ones; in almost all problems, alternative approaches and solutions are possible that are counted as correct.

SECTION A

(1 hour 50 minutes + 10 minutes for reading)

Answer **ALL** questions from this section (questions **1-4**).

Each question in this section bears 15 marks

Question 1. (15 marks) Let the regression equation be:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t \qquad t = 1, 2, ..., T .$$

(a) Outline how you would test the following restrictions

a.1) $\beta_3 = 0$,

a.2) $\beta_2 = 1$

a.3) $\beta_3 = 0$ and $\beta_2 = 1$ simultaneously.

Specify the assumptions required for these tests. How to estimate the relevant equations needed for the test?

<u>Solution</u>

a.1) This is a standard two-tail *t* test of the form: $t = \dfrac{\hat{\beta}_3}{\text{se}_{\hat{\beta}_3}}$.

The degrees of freedom for the *t* test is $T-4$.

a.2) Now we use statistic $t = \dfrac{\hat{\beta}_2 - 1}{\text{se}_{\hat{\beta}_2}}$ with the same degrees of freedom and the same critical values.

a.3) To test two restrictions simultaneously the F-test should be used. First run the original unrestricted equation

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t$$

to get residual sum of squares RSS_U. Then write down the restricted version of it

$$Y_t = \beta_1 + 1 \cdot X_{2t} + 0 \cdot X_{3t} + \beta_4 X_{4t} + u_t$$

To estimate equation we should transform it into the form

$$Y_t - X_{2t} = \beta_1 + \beta_4 X_{4t} + u_t$$

and then run simple linear regression to obtain restricted residual sum of squares RSS_R. Then apply F-test $F = \dfrac{(RSS_R^2 - RSS_U^2)/2}{RSS_U^2/(T-4)}$ with 2 and $T-4$ degrees of freedom.

Assumptions: u_t is distributed normally and standard errors are valid what can be doubted, since time series are usually characterized by the presence of autocorrelation.

(b) Outline how you would test the following restrictions

b.1) $\beta_2 = \beta_3 = \beta_4$ simultaneously against NOT $\beta_2 = \beta_3 = \beta_4$

b.2) $\beta_2 + \beta_3 + \beta_4 = 1$ against $\beta_2 + \beta_3 + \beta_4 \neq 1$,

Indicate test statistics. Specify the numbers of degrees of freedom. How to estimate the relevant equations needed for the tests?

Solution

b.1) First run unrestricted unrestricted equation

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t$$

to get residual sum of squares RSS_U. Then write down the restricted version of it

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_2 X_{3t} + \beta_2 X_{4t} + u_t .$$

Then run simple linear regression $Y_t = \beta_1 + \beta_2 (X_{2t} + X_{3t} + X_{4t}) + u_t$ to obtain restricted residual sum of squares RSS_R. Then apply the same F-test $F = \dfrac{(RSS_R^2 - RSS_U^2)/2}{RSS_U^2/(T-4)}$ with the same 2 and $T-4$ degrees of freedom.

b.2) First run unrestricted unrestricted equation $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t$ to get residual sum of squares RSS_U. Restricted equation one can write in the form $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + (1 - \beta_2 - \beta_3) X_{4t} + u_t$. To estimate equation we should transform it into the form

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + 1 \cdot X_{4t} - \beta_2 X_{4t} - \beta_3 X_{4t} + u_t \text{ or}$$

$$Y_t - X_{4t} = \beta_1 + \beta_2 (X_{2t} - X_{4t}) + \beta_3 (X_{3t} - X_{4t}) + u_t .$$

Then run restricted regression to obtain residual sum of squares RSS_R. Then apply F-test $F = \dfrac{(RSS_R^2 - RSS_U^2)/1}{RSS_U^2/(T-4)}$ with 1 and $T-4$ degrees of freedom.

(c) Outline how you would test the following restrictions

$\beta_2 + \beta_3 + \beta_4 = 1$ against $\beta_2 + \beta_3 + \beta_4 > 1$,

Indicate test statistic. Specify the number of degrees of freedom. How to estimate the relevant equations needed for the test?

Solution

Here F-test is not applicable as the it is always two sided while we have here one sided alternative. So we have to use reparametrisation and then apply t-test. To do this we need to use left side of restriction $\beta_2 + \beta_3 + \beta_4 - 1 = 0$ in explicit form as the coefficient of regression. So let us add (and subtract simultaneously) to equation

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t$$

some terms $\beta_3 X_{2t} + \beta_4 X_{2t} - 1 \cdot X_{2t}$:

$$Y_t = \beta_1 + \beta_2 X_{2t} + (\beta_3 X_{2t} + \beta_4 X_{2t} - 1 \cdot X_{2t}) - (\beta_3 X_{2t} + \beta_4 X_{2t} - 1 \cdot X_{2t}) + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t \text{ or}$$

$$Y_t = \beta_1 + (\beta_2 + \beta_3 + \beta_4 - 1) X_{2t} + 1 \cdot X_{2t} + \beta_3 (X_{3t} - X_{2t}) + \beta_4 (X_{4t} - X_{2t}) + u_t \text{ or}$$

$$Y_t - X_{2t} = \beta_1 + (\beta_2 + \beta_3 + \beta_4 - 1) X_{2t} + \beta_3 (X_{3t} - X_{2t}) + \beta_4 (X_{4t} - X_{2t}) + u_t$$

Now we can run the regression and apply one-sided t-test to the coefficient of the regressor X_{2t} with $T-4$ degrees of freedom.

Question 2.

(15 marks) March 18, 2018 in Russia held the election of the President of the Russian Federation. The 3rd grade student of ICEF decided to collect data from the official reports of 87 regional election commissions. He built the equations expressing the dependence of the votes for each candidate in turn (BA_i - Baburin share, GR_i - Grudinin share, PU_i - Putin share, SO_i - Sobchak share, SU_i - Suraykin share, TI_i - Titov share, YA_i - Yavlinsky share, ZH_i - Zhirinovsky share) on the voters' activity, meaning the proportion of those who voted for a certain candidate in the total number of registered voters ($PART_i$ - participation). For example the equation for Sobchak:

$$\underset{(0.005)}{SO_i = 0.033} - \underset{(0.008)}{0.026 PART_i} \qquad R^2 = 0.119 \tag{1}$$

Then he added additional variable PU_i – the share of the votes for Putin

$$\underset{(0.009)}{SO_i = 0.039} - \underset{(0.012)}{0.018 PART_i} - \underset{(0.017)}{0.011 PU_{ii}} \qquad R^2 = 0.127 \tag{2}$$

(a) Comment on the differences in R^2 for each equation. Are the variables in equations (1) and (2) significant taken by one and together? What is the difference in interpretation of the coefficients of the common variable $PART_i$ in (1) and (2)? The student noticed that all slope coefficients in (1) and (2) have minus sign. Help him to find some explanation to this for each coefficient in turn.

Disappointed by the insignificance of the slope coefficients in (2) the student decided to add all other alternative candidates to equation

$$SO_i = \beta_1 + \beta_2 PART_i + \beta_3 BA_i + \beta_4 GR_i + \beta_5 PU_i + \beta_6 SU_i + \beta_7 TI_i + \beta_8 YA_i + \beta_9 ZH_i + u_i$$

but student's friend said that he can guess the values of almost all the coefficients without evaluating the equation. Try to do the same and explain?

Solution

The determination coefficient R^2 in equations (2) is slightly bigger than in (1) what is naturally as new variable is added. The slope coefficient in equation (1) is significant as $t = \frac{-0.026}{0.008} = -3.25$ with $df = 87 - 2 = 85$, while $t_{crit}(80; 1\%) = 2.639$. Both slope coefficients in equation (2) are insignificant $t = \frac{-0.018}{0.012} = -1.5$ and $t = \frac{-0.011}{-0.017} = -0.647$ while $t_{crit}(80; 5\%) = 1.99$. The variables $PART_i$ and PU_i taken together are significant as $F = \frac{0.127/2}{(1-0.127)/(87-3)} = 6.11$ while $F_{crit}(2, 80; 1\%) = 4.88$.

While coefficient -0.026 in (1) shows marginal effect of participation on Sobchak share of votes, the coefficient $-0.018 > -0.026$ shows marginal effect of participation on Sobchak share keeping Putin share constant (it is clear that with the restraint of the main competitor, the odds of Sobchak increases).

Negative sign of the coefficient of PU_i is quite expect, since Putin is an alternative candidate for Sobchak.

Negative sign of the coefficient of $PART_i$ could be probably explained by the fact that increasing participation of voters increases the shares of all candidates but mainly the share of the favorite of presidential election. We should have in mind that the coefficient of this variable in equation (1) is also under the omitted variable bias.

The suggested equation express the fact that the shares of all candidates in the sum give unity, so the intercept of this equation is equal to 1, while slope coefficients of all candidates are equal to -1, the value of the coefficient of $PART_i$ is unpredictable.

(b) Then he decided to investigate how regional factor influence the share of votes for Putin. He divides all regions into 4 categories R1 (central regions with big cities such as Moscow and St.Petersburg), R2 (North, Siberia and Far East), R3 (Caucasus and South), R4 (Crimea and Sevastopol). Then he estimated the following equations:

$$\underset{(0.033)}{PU_i = 0.420} + \underset{(0.048)}{0.517 PART_i} \qquad R^2 = 0.582 \tag{3}$$

$$\underset{(0.030)}{PU_i = 0.461} + \underset{(0.044)}{0.475 PART_i} - \underset{(0.008)}{0.033 R2_i} + \underset{(0.013)}{0.003 R3_i} + \underset{(0.021)}{0.089 R4_i} \qquad R^2 = 0.731 \tag{4}$$

A friend of our student looked at the models and pointed out that model (4) could not be good as a dummy for central regions with big cities was omitted. Comment upon this criticism.

Give interpretation to the coefficients of equations (3) and (4). Is regional factor significant based on comparison of these equations? What test is applicable here? Chow test applied to equation (3) gives F statistic equal to 9.28, compare this value with the value from your test and comment.

Solution

The student should not follow the advice of his friend, since in this case a dummy variables trap will arise. The coefficient $+0.517$ in (3) shows positive marginal effect of participation on Putin results (while for Sobchak it is negative). The constant term $+0.420$ in (3) cannot be interpreted as it is obvious that al observation of the variable 'participation' are far from zero.

In equation (4) the intercept $+0.461$ relates to the central region while for example for North, Siberia and Far East it is less by 0.033, $+0.475$ is marginal effect of participation common for all regions.

Regional factor is significant as $F = \dfrac{(0.731 - 0.583)/3}{(1 - 0.731)/(87 - 5)} = 15.04$ while $F_{crit}(3, 80; 1\%) = 4.04$.

Chow test gives different result as it is equivalent to the F-test for the joint significance of all dummies only when also slope dummies $PART_i \cdot R2$, $PART_i \cdot R3$, $PART_i \cdot R4$, are included.

(c) During the discussion of the results of student's study his scientific adviser mentioned on the possible endogeneity in the equation 3. So the student evaluated auxiliary equation

$$\hat{PART_i} = \hat{\beta}_0 + \hat{\beta}_1 VOTES_i + \hat{\beta}_1 SOBCHAK_i + \hat{\beta}_1 ZHIRIN_i \tag{5}$$

using $VOTES_i$ (total number of votes), $SOBCHAK_i$ (number of votes for Sobchak), $ZHIRIN_i$ (number of votes for Zhirinovsky) as the instruments for $PART_i$, and then added the residuals of this equation E_i in equation (3)

$$\underset{(0.050)}{PUT_i = 0.287} + \underset{(0.073)}{0.712 PART_i} - \underset{(0.092)}{0.314 E_i} \qquad R^2 = 0.633 \tag{6}$$

Explain what is endogeneity. Explain the procedure above, indicating null hypothesis and test statistic, and comment the results of the test. Discuss used set of instruments to conclude under what conditions the used test is valid?

Solution

Endogenicity means that the variable used as an independent variable is partially determined by the dependent variable of the equation, and therefore correlates with the random term of the equation. This violates Gauss-Markov conditions and makes estimators and tests invalid. The student did Hausmann test for endogeneity in Davidson-McKinnon form.

In the original Durbin-Wu-Hausmann test the test statistic summarizes the differences between the OLS and IV estimates of the parameters and a chi-squared test is used to test whether this summary statistic is significant. If there is no measurement error, IV and OLS are both consistent but OLS is to be preferred because it is more efficient. If there is measurement error, IV is consistent but OLS is not.

The version of this test proposed by Davidson and MacKinnon) includes two stages.

1) we regress suspect variable on all exogenous variables and instruments and retrieve the residuals (auxiliary regression).

2) then we re-estimate main regression including the residuals from the first (auxiliary) regression as additional regressors.
If the OLS estimates are consistent, then the coefficient on the first stage residuals should not be significantly different from zero.
So the significance of the residuals could be considered as a sign of possible endogeneity of suspect explanatory variable.

Following this algorithm the student estimated potentially endogenous variable $PART_i$ on the set of available instruments

$$\hat{PART}_i = \hat{\beta}_0 + \hat{\beta}_1 VOTES_i + \hat{\beta}_1 SOBCHAK_i + \hat{\beta}_1 ZHIRIN_i$$

saved the residuals E_i and then added the residuals of this auxiliary equation in equation (3). It remains to verify the coefficient with the added residuals using the usual t-statistics: $t = \frac{-0.314}{0.092} = 3.41$ while $t_{crit}(80; 1\%) = 2.64$ so null hypothesis is rejected.
The test is valid only if used variables can be considered as valid instruments, but there are serious doubts about this. With a high probability, these variables are themselves endogenous, and then cannot be considered as tools. The student should look for tools among the exogenous variables.

Question 3.
The student is interested in the influence of the price level and income on the consumption of soft drinks SD_{it}. Suppose that the time series C_t (Coca-cola), P_t (Pepsi), S_t (Sprite), U_t (7-up), M_t (Myrinda), F_t (Fanta) for 1988-2017 form the panel (as observations for particular units of some general type of soft drinks (SD_{it})). DPI_t is the disposable personal income and $PRSD_{it}$ is the relative price index for corresponding SD_{it} .Let the model under investigation be

$$\log(SD_{it}) = \beta_1 + \beta_2 \log(DPI_t) + \beta_3 \log(PRSD_{it}) + u_{it}\,.$$

(a) What are the advantages of panel data analysis comparing to cross-section regression and time series analysis? What are typical problems associated with the panel data models? The student decided to use the approach based on the fixed effects panel data model. Help the student to understand what is LSDV method and how to use it for estimation of panel data model under consideration. Explain clearly what are fixed effects.

<u>Solution</u>
What are advantages of panel data analysis comparing to cross-section regression and time series analysis?
Larger data set. If in time series data T observations are available, with panel data T*n observations are available, where n is the number of units and T is the number of periods.
Unobserved heterogeneity problem (more about it later) is eliminated or mitigated.
Dynamics can be explored better (compared to cross-section). Although with cross-section one could investigate dynamics by asking retrospective questions, it is not very reliable, as people forget details over time.
Panel data are often of higher quality. For example, national surveys are usually rather well designed and well organized.

What are typical problems associated with the panel data models?
1. Cost of data collection
2. Issue of modeling unobserved heterogeneity

The main question in the panel data analysis is the problem of the origin of unobserved heterogeneity $\sum_{p=1}^{s} \gamma_p Z_{pi}$ $= \alpha_i$ in the model of the type

$$Y_{it} = \beta_0 + \sum_{j=1}^{k} \beta_j X_{jit} + \sum_{p=1}^{s} \gamma_p Z_{pi} + \varepsilon_{it} \qquad (1).$$

Fixed effect approach assigns it to the fixed characteristics of individual elements (here to the types of soft drinks), while random effect approach assigns it to the random factors.

In the **Least Squares Dummy Variable (LSDV) fixed effect method** the unobserved effect is brought explicitly into the model. A set of dummy variables D_i is defined, where D_i is equal to 1 in the case of an observation relating to an individual *i* and 0 otherwise. The model can be written as

$$Y_{it} = \sum_{j=2}^{K} \beta_j X_{jit} + \sum_{i=1}^{n} \alpha_i D_i + u_{it}. \qquad \text{(vi)}$$

The unobserved effect is now being treated as the coefficient of the specific individual *i*.

The term $\alpha_i D_i$ represents a fixed effect on the dependent variable Y_i for individual *i*. We use the model without intercept otherwise we will fall into the dummy variable trap. Entering the details, we note that the method calculates its own intercept as an average of all coefficients for dummies, and represents fixed effects as the deviations of these coefficients from the calculated mean. The method is based on the assumption that the elasticities in income and prices are equal for all soft drinks, and heterogeneity is reflected only in the intercept.

It should be noted that inclusion of a set of dummies leads to losing of additional degrees of freedom, so estimators become less efficient.

(b) The model above is based on the assumption that the elasticities of different goods are the same. Suggest how to test this assumption. Give some details: indicate corresponding test and the data needed for it, null hypothesis, distribution of the test statistic, the number of degrees of freedom and the decision rule.

Solution

We start from LSDV fixed effects model of the type

$$\log(SD_{it}) = \beta_2 \log(DPI_t) + \beta_3 \log(PRSD_{it}) + \sum_{i=1}^{6} \gamma_i D_i + u_{it}. \qquad t = 1, 2, ..., 30 \quad (1)$$

where D_i are dummies corresponding different drinks under consideration. To take into account that the elasticities of DPI_t and $PRSD_{it}$ could be different, it is possible to run instead 6 different regressions for 6 drinks ($i = 1, 2, ..., 6$),

$$\log(SD_{it}) = \gamma_1 + \gamma_2 \log(DPI_t) + \gamma_3 \log(PRSD_{it}) + u_{it}$$

and then perform an F-test using the values of RSS' of evaluated models:

$$F = \frac{(RSS_{LSDV} - \sum_{i=1}^{6} RSS_i)/(6\cdot 3-(2+6))}{\sum_{i=1}^{6} RSS_i/(30-6\cdot 3)} = \frac{(RSS_{LSDV} - \sum_{i=1}^{6} RSS_i)/10}{\sum_{i=1}^{6} RSS_i/12}.$$

Here RSS_{LSDV} - the value of RSS for LSDV fixed effect model, $\sum_{i=1}^{6} RSS_i$ - sum of RSS' for separate regressions for different goods, the rule for degrees of freedom follows general principles of F-test and is clear from the formula above. If H_0 is not rejected there is no significant differences between elasticities.

(c) What other methods within the fixed effects panel data model can the student apply? What are their comparative advantages and disadvantages? Could the researcher have used a random effects regression in the present case? How to choose between random effects and fixed effects panel model?

Solution

What other methods within the fixed effects panel data model can the student apply? **(1 point)**

1. Within groups method

2. First differences method

What are their comparative advantages and disadvantages? **(1 point)**

1. All three methods (LSDV, Within groups and First difference) leads to losing $n-1=4-1=3$ degrees of freedom.
2. Autocorrelation problem may appear if researcher uses First difference method (and in a weakened form in the Within groups method), which may even be useful, since the resulting autocorrelation of the MA (1) type is negative, which, superimposed on the positive autocorrelation in the original data, can solve or alleviate the problem of autocorrelation as a whole.
3. Variation of new explanatory variables is smaller, thus precision of estimates decreases
4. All variables are constant in time

Could the researcher have used a random effects regression in the present case? ***(1 point)***

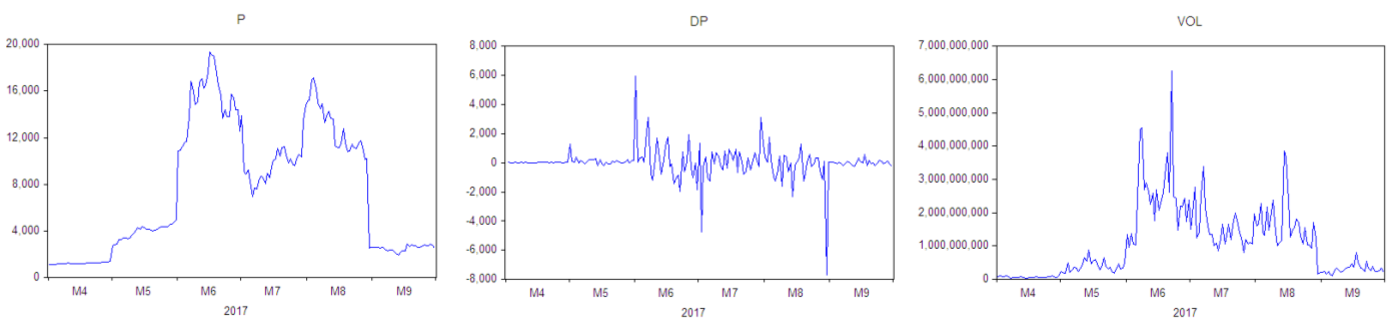
The student is recommended to use Darbin-Wu-Hausman (DWH) test to choose between fixed and random effects. It is standard for majority of econometric computer programs and is based on using chi-square statistics with degrees of freedom equal to the number of variables in the equation under consideration (2 in our case) (as it compares estimates of coefficients obtained by two alternative models). Under H_0 that there is no difference between coefficients obtained by two alternative models – fixed and random panel models (which means that unobserved heterogeneity α_i as a part of disturbance term, is not correlated with DPI_t and $PRSD_{it}$) both fixed effect and random effect models provide us with consistent estimates. We choose in this case random effect models as it retains in disturbance term all unobserved heterogeneity, there is no reduction of degrees of freedom typical for fixed effects models.

If H_0 is rejected, so there are essential differences between coefficients obtained using fixed and random effects models, we choose fixed effects model, because rejecting of H_0 means that main assumption of independence of the disturbance term from regressors is violated so using random effects model we are under risk of getting inconsistent estimates of parameters. So we have to suffice the fixed effects model that always gives consistent estimates.

It should be noted that random effect model could not be applied in this case from economic point of view as for this model first condition of sample being drawn randomly from the population, but Fanta belongs to Coca Cola company, while Myrinda, 7up and Sprite belong to PepsiCo, thus these soft drinks could not be considered separately, because their pricing policies are not independent.

Question 4.

(15 marks) Using daily data for the period from 2017-04-01 to 2017-09-30 (183 observations), a student tries to investigate the behavior of bitcoin prices P_t (in dollars), their differences DP_t and corresponding dynamics of the volume of trade VOL_t (see the corresponding graphs below). For time trend $t=0$ at $4/01/2017$.



Suspecting that some of these time series might be non-stationary she runs some Dickey-Fuller equations to test the series for stationarity:

$$\hat{\Delta P_t} = 160.58 - 0.02 P_{t-1} \qquad R^2 = 0.01, \qquad \textbf{(1)}$$
$$(134.00) \quad (0.014)$$

$$\Delta \hat{DP_t} = -0.97 DP_{t-1} \qquad R^2 = 0.487 \qquad \textbf{(2)}$$
$$(0.075)$$

$$\Delta \hat{VOL_t} = 1.48 \cdot 10^8 - 0.144 VOL_{t-1} - 0.224 \Delta(VOL_{t-1}) + 91320.24t \qquad R^2 = 0.14 \qquad \textbf{(3)}$$
$$(871445.3) \quad (0.045) \qquad (0.073) \qquad (871445.3)$$

$$\Delta \hat{VOL_t} = 1.55 \cdot 10^8 - 0.143 VOL_{t-1} - 0.224 \Delta(VOL_{t-1}) \qquad R^2 = 0.14 \qquad \textbf{(4)}$$
$$(65210866) \quad (0.044) \qquad (0.073)$$

(a) Looking at the graphs what are your initial guesses about stationarity of the given time series? Describe briefly how to use Dickey-Fuller equations to test time series under consideration for stationarity. For equations (1) - (4) perform the (Augmented) Dickey-Fuller t-test, each time clearly indicating the null hypothesis, the test statistic and critical values. Simultaneously explain how to find the critical values in the presence of a time trend? How to test the presense of time trend in time series? Compare the results of equations (3) and (4). Comment on the results.

Solution

It would be good to mention here what stationarity means (mean, variance, covariance - independent of time), then to comment on the graphs, the 1st one– cannot probably be stationary (the mean is changing, probably some structural break in the middle), the 2nd graph– quite random, the mean changes not that much, variance is relatively constant, so it looks more or less stationary (by the way, if it's stationary it implies that $P_t \sim I(1)$); the 3rd graph - during M6-M9 - probably there is a trend, but more likely that the shock/disturbance is just fading over time – in general looks like VOL_t is stationary but probably not – need to perform tests.

For finite sample to test for stationarity we should evaluate Dickey-Fuller t-statistic using information from Dickey-Fuller equation (for example $P_t = \beta_1 + \beta_2 P_{t-1} + u_t$ => $\Delta P_t = \beta_1 + (\beta_2 - 1)P_{t-1} + u_t$ (1)) and then use special tables for this statistic as under null hypothesis $H_0: \beta_2 - 1 = 0$ time series is nonstationary, so conventional t-test is invalid.

For example for equation (1) $t = \dfrac{-0.02}{0.014} = -1.43$. Using DF t-statistic tables for 183 observation (rounding up 150) we find $t(ADF)_{crit}(No\,trend, 5\%, 150\,obs) = -2.88$. As $-2.88 < -1.43$ the null hypothesis of nonstationarity of the time series P_t is not rejected.

Similarly for equation (2) $t = \dfrac{-0.97}{0.075} = -12.93$ while $t(ADF)_{crit}(No\,trend, 1\%, 200\,obs) = -3.46$, so $-12.93 < -3.46$ and null hypothesis of nonstationarity of the time series DP_t is rejected.

The obtained conclusions generally correspond to the first two graphs above.

Equations (3) and (4) have a more complex form

$$\Delta VOL_t = \beta_1 + (\beta_2 + \beta_3 - 1)VOL_{t-1} - \beta_3 \Delta VOL_{t-1} + \beta_4 t + u_t$$

and

$$\Delta VOL_t = \beta_1 + (\beta_2 + \beta_3 - 1)VOL_{t-1} - \beta_3 \Delta VOL_{t-1} + u_t$$

correspondingly, so the null hypothesis is now $H_0: (\beta_2 + \beta_3 - 1) = 0$, but from practical side almost everything is similar.

For equation (3) $t = \dfrac{-0.144}{0.045} = -3.2$ while $t(ADF)_{crit}(Trend\ in\ model, 5\%, 150\,obs) = -3.44$, so $-3.44 < -3.2$ and null hypothesis of nonstationarity of the time series VOL_t cannot be rejected at 5%.

(One can mention here that practically we test whether the time series VOL_t is difference stationary (under the null) against the alternative of VOL_t being trend stationary (and so we've got no confirmation of trend stationarity here).

The test allows also to test the presence of time trend, using using the usual t-test tables: $t = \dfrac{91320.24}{871445.3} = 0.105$ while $t_{crit}(5\%, 18) = 1.976$ so time trend is insignificant. It can be also seen on the graph above.

So the variable t is redundant and can be excluded from equation to make estimators more efficient. In fact for equation (4) t-statistic is almost the same $t = \dfrac{-0.143}{0.044} = -3.25$ but now we use critical value without time trend $t(ADF)_{crit}(No\,trend, 5\%, 200\,obs) = -2.88$, so $-3.25 < -2.88$ and null hypothesis of nonstatonarity of the time series VOL_t can be rejected at 5% so the time series VOL_t is stationary.

(b) Now for equations (1-4) carry out the Dickey-Fuller test using the scaled estimator of the slope coefficient $T(\hat{\beta}_2 - 1)$ to test series P_t, DP_t and VOL_t for nonstationarity where β_2 is the slope coefficient of the autoregression $Y_t = \beta_2 Y_{t-1} + u_t$. Indicate in each case the null hypothesis and used critical values. Do the results of these tests coincide with your conclusions based on t-tests? Comment the meaning of $T(\hat{\beta}_2 - 1)$ statistic and explain why the difference $(\hat{\beta}_2 - 1)$ should be multiplied by T rather than by $\sqrt{T}$.

Solution

In addition to the ADF t-test Dickey and Fuller proposed another test for nonstationarity involving a scaled estimator $T(\hat{\beta}_2 - 1)$ of the slope coefficient of the autoregressive process $Y_t = \beta_2 Y_{t-1} + u_t$. If $|\beta_2| < 1$ then $\sqrt{T}(\hat{\beta}_2 - \beta_2)$ has limiting distribution $\sqrt{T}(\hat{\beta}_2 - \beta_2) \to N(0, 1-\beta^2)$ (multiplication by $\sqrt{T}$ is sufficient to prevent the variance from tending to zero. But in case of $\beta_2 = 1$ (null hypothesis of random walk) the distribution is contracting to the true value faster than the standard rate so the multiplying by $\sqrt{T}$ is not enough to prevent the distribution collapsing to a skipe (the estimator $\hat{\beta}_2$ in this case is called superconsistent), and we should multiply by T to get statistic $T(\hat{\beta}_2 - 1)$. For finite sample to test for stationarity we should evaluate Dickey-Fuller $T(\hat{\beta}_2 - 1)$ statistic using information from Dickey-Fuller equation and then use special tables for this statistic.

For example equation (1) has the form $\Delta P_t = \beta_1 + (\beta_2 - 1)P_{t-1} + u_t$, so $(\hat{\beta}_2 - 1) = -0.02$ and $T = 183$, so $T(\hat{\beta}_2 - 1) = 183 \cdot (-0.02) = -3.66$. Using DF tables we find $T(\hat{\beta}_2 - 1)_{crit}(No\,trend, 5\%, 150\,obs) = -13.72$. As $-13.72 < -3.66 < 0$ the null hypothesis of nonstationarity of the time series P_t is not rejected.

Similarly for equation (2) $T(\hat{\beta}_2 - 1) = 183 \cdot (-0.97) = -177.5$ while $T(\hat{\beta}_2 - 1)_{crit}(No\,trend, 1\%, 200\,obs) = -20.02$, so $-177.5 < -20.02$ and null hypothesis of nonstationarity of the time series DP_t is rejected.

For equation (3) $183 \cdot (-0.144) = -26.35$ while $T(\hat{\beta}_2 - 1)_{crit}(Trend\ in\ model, 5\%, 200\,obs) = -21.06$, so $-26.35 < -21.06$ and null hypothesis of nonstationarity of the time series VOL_t is rejected.

For equation (4) $183 \cdot (-0.143) = -26.17$ while $T(\hat{\beta}_2 - 1)_{crit}(No\,trend, 1\%, 200\,obs) = -20.02$, so $-26.17 < -20.02$ and null hypothesis of nonstationarity of the time series VOL_t is rejected.

NOTE: To be more precise in case of additional lags the statistic is slightly different: it includes the division by $(1 - coefficients\ of\ lagged\ differences)$: for example for equation (3) $\dfrac{183}{1+0.224} \cdot (-0.144) = -21.53$ while $T(\hat{\beta}_2 - 1)_{crit}(Trend\ in\ model, 5\%, 200\,obs) = -21.06$, so $-21.53 < -21.06$ and null hypothesis of nonstationarity of the time series VOL_t is rejected. For equation (4) $\frac{183}{1+0.224}\cdot(-0.143) = -21.34$ while $T(\hat{\beta}_2 - 1)_{crit}(No\,trend, 1\%, 200\,obs) = -20.02$, so $-21.34 < -20.02$ and null hypothesis of nonstationarity of the time series VOL_t is rejected. But students are not required to use this amendment, as the corresponding theory lies outside our introductory course.

The results generally coincide with those obtained in a), but unlike insignificant t-statistic for equation (3) the statistic $T(\hat{\beta}_2 - 1)$ gives significant result.

(c) The student obtained for equation (1)

$$\Delta P_t = \beta_1 + (\beta_2 - 1)P_{t-1} + u_t,$$

the value of F-statistics for testing simultaneously two restrictions $\beta_1 = 0, \beta_2 = 1$: $F = 1.00$. How to use this information to conduct Dickey-Fuller F-test for the nonstationarity of the P_t? Following this approach describe how you can investigate also eqations (2-4) to test the series DP_t and VOL_t for the nonstationarity using Dickey-Fuller F-test: for each case indicate theoretical equation, restriction(s), and the rule for choosing the critical value of ADF F-statistic from the appropriate table. Indicate also in each case the random processes which the test allows to discriminate. What are comparative advantages and disadvantages of three different ADF tests for nonstationarity?

Solution

There is the third form of Dickey-Fuller test, based on F-statistic. According the restrictions above it allows to discriminate the autoregressive process with intercept $P_t = \beta_1 + \beta_2 P_{t-1} + u_t$ and random walk $P_t = P_{t-1} + u_t$. Comparing the value $F = 1.00$ with $F(ADF)_{crit}(No\,trend, 5\%, 150\,obs) = 4.67$ we can conclude that null hypothesis of nonstationarity ($H_0: \beta_2 - 1 = 0, \beta_1 = 0$) cannot be rejected.

For equation (2) this approach adds nothing as for equation

$$\Delta DP_t = (\beta_2 - 1)DP_{t-1} + u_t$$

the value of F-statistic for restriction $\beta_2 = 1$ is exactly the square of t-statistic of coefficient $\beta_2 - 1$ in equation (2).

For equation (3) containing intercept and time trend

$$\Delta VOL_t = \beta_1 + (\beta_2 - 1)VOL_{t-1} + \beta_3 \Delta VOL_{t-1} + \beta_4 t + u_t$$

β_1 is now unconstrained (as well as β_3) because a process cannot combine a random walk with drift and time trend so we again have two restrictions that could be tested with an F-test: $\beta_2 = 1, \beta_4 = 0$, corresponding F-statistics should be compared with $F(ADF)_{crit}(Trend\ in\ model, \alpha\%)$, for example for 5% it is $F(ADF)_{crit}(Trend\ in\ model, 5\%, 150\,obs) = 6.395$. This test allows discriminating the process including determistic trend (so it cannot be called stationary but trend is minor problem as it can be removing by detrending) and the process including random walk with drift $VOL_t = \beta_1 + VOL_{t-1} + \beta_3 \Delta VOL_{t-1} + u_t$

We have argued in (a) that time trend is in fact insignificant so it can be excluded from equation and we come to equation (4) without trend,

$$\Delta VOL_t = \beta_1 + (\beta_2 - 1)VOL_{t-1} + \beta_3 \Delta VOL_{t-1} + u_t$$

Here again we have two restrictions $\beta_1 = 0, \beta_2 = 1$ so after obtaining corresponding F-statistic we can use the appropriate critical values $F(ADF)_{crit}(No\,trend, \alpha\%)$ to discriminate between stationary process and the process including random walk without drift

$$VOL_t = VOL_{t-1} + \beta_3 \Delta VOL_{t-1} + u_t$$

When β_2 tends to 1, the power of Dickey-Fuller tests decreases rapidly, there is a difficulty of discriminating nonstationary process and stationary process that is highly autoregressive. Comparing all three Dickey-Fuller tests (based on t-statistic, F-statistic and $T(\hat{\beta}_2 - 1)$ statistic) we shoud keep in mind that the one based on $T(\hat{\beta}_2 - 1)$ statistic has the biggest power and one based on F-statistic has the smallest power. Nevertheless the ADF t-test remains most popular as it can be evaluated using available software.

SECTION B

(1 hour 10 minutes + 10 minutes for reading)

Answer **TWO** questions from this section (questions **5-7**).

Each answered question in this section bears 20 marks.

Question 5.

Suppose the model $Y_i = \beta_1 + \beta_2 X_i + u_i,\ \ i = 1, \ldots, n$ satisfies all assumptions of the model A (Gauss-Markov conditions). Let the values of X_i are supposed to be non-stochastic. The researcher wrongly believes that $\beta_1 = 0$ and so uses OLS estimator $\tilde{\beta}_2$ for the model $Y_i = \beta_2 X_i + u_i,\ \ i = 1, \ldots, n$ instead of OLS estimator $\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$ for the model $Y_i = \beta_1 + \beta_2 X_i + u_i,\ \ i = 1, \ldots, n$ where $x_i = X_i - \overline{X}$, $y_i = Y_i - \overline{Y}$.

(a) Show that OLS estimator $\tilde{\beta}_2$ for the model $Y_i = \beta_2 X_i + u_i,\ \ i = 1, \ldots, n$ is $\tilde{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2}$.

Solution

Under assumption $\beta_1 = 0$ OLS is minimization of a function $F(\beta_2) = \sum (Y_i - \beta_2 X_i)^2$. First order condition for this is $\frac{d}{d\beta_2} \sum (Y_i - \beta_2 X_i)^2 = 0$ or $\sum (Y_i - \beta_2 X_i) X_i = 0$ that gives $\tilde{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2}$

(b) If the assumption $\beta_1 = 0$ is not correct show that estimator $\tilde{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2}$ is biased and find the bias.

Indicate the case when despite $\beta_1 \neq 0$ the estimator $\tilde{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2}$ is nevertheless unbiased.

Solution

$E\tilde{\beta}_2 = E\frac{\sum X_i Y_i}{\sum X_i^2} = \frac{\sum X_i EY_i}{\sum X_i^2} = \frac{\sum X_i E(\beta_1 + \beta_2 X_i + u_i)}{\sum X_i^2} = \frac{\sum X_i (\beta_1 + \beta_2 X_i + Eu_i)}{\sum X_i^2}$. According to G.M.C. $Eu_i = 0$ so $E\tilde{\beta}_2 = \frac{\sum X_i(\beta_1 + \beta_2 X_i)}{\sum X_i^2} = \frac{\sum \beta_1 X_i + \sum \beta_2 X_i^2}{\sum X_i^2} = \beta_2 + \beta_1 \frac{\sum X_i}{\sum X_i^2}$. The bias is $\beta_1 \frac{\sum X_i}{\sum X_i^2}$.

The bias disappears if $\sum X_i = 0$ under assumption $\beta_1 \neq 0$.

(c) Show that the variance of $\tilde{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2}$ is given by expression $\operatorname{var}(\tilde{\beta}_2) = \frac{\sigma^2}{\sum X_i^2}$, where $\sigma^2 = \operatorname{var}(u_i)$.

Solution

$$\operatorname{var}(\tilde{\beta}_2)=\operatorname{var}\left(\frac{\sum_{t=1}^{T}X_tY_t}{\sum_{t=1}^{T}X_t^2}\right)=\frac{\operatorname{var}\left(\sum_{t=1}^{T}X_tY_t\right)}{\left(\sum_{t=1}^{T}X_t^2\right)^2}=\frac{\sum_{t=1}^{T}(\operatorname{var}X_tY_t)}{\left(\sum_{t=1}^{T}X_t^2\right)^2}=\frac{\sum_{t=1}^{T}X_t^2(\operatorname{var}Y_t)}{\left(\sum_{t=1}^{T}X_t^2\right)^2}=\frac{\sum_{t=1}^{T}X_t^2\sigma^2}{\left(\sum_{t=1}^{T}X_t^2\right)^2}=\frac{\sigma^2}{\sum_{t=1}^{T}X_t^2}$$

(d) Show that if assumption $\beta_1=0$ is true the estimator $\hat{\beta}_2=\frac{\sum x_iy_i}{\sum x_i^2}$ is still unbiased but inefficient.

(Hint: compare the variances of two estimators $\tilde{\beta}_2=\frac{\sum X_iY_i}{\sum X_i^2}$ and $\hat{\beta}_2=\frac{\sum x_iy_i}{\sum x_i^2}$).

Solution

As it is known $\hat{\beta}_2=\frac{\sum x_iy_i}{\sum x_i^2}=\beta_2+\frac{\sum x_iu_i}{\sum x_i^2}$. Under assumption $\beta_1=0$ $E\hat{\beta}_2=\beta_2+\frac{\sum x_iEu_i}{\sum x_i^2}=\beta_2$, as $Eu_i=0$, so the estimator still remains unbiased.

It is known that $\operatorname{var}(\hat{\beta}_2)=\frac{\sigma^2}{\sum x_i^2}$ and we proved in **(c)** that $\operatorname{var}(\tilde{\beta}_2)=\frac{\sigma^2}{\sum X_i^2}$ ..So we need to compare both variances. Note that $\sum x_i^2=\sum(X_i-\overline{X})^2=\sum X_i^2-2\overline{X}\sum X_i+\sum(\overline{X})^2=\sum X_i^2-n(\overline{X})^2$ and obviously $\sum X_i^2-n\sum(\overline{X})^2\le\sum X_i^2$ so $\operatorname{var}(\hat{\beta}_2)=\frac{\sigma^2}{\sum x_i^2}=\frac{\sigma^2}{\sum X_i^2-n(\overline{X})^2}\ge\frac{\sigma^2}{\sum X_i^2}=\operatorname{var}(\tilde{\beta}_2)$.

That means that if assumption $\beta_1=0$ is correct $\hat{\beta}_2$ has bigger variance and so is no more efficient estimator of a slope.

Question 6.

1. Consider a model

$$Q_t=\alpha+\beta P_t^*+\gamma Z_t+u_t;\ t=1,2,...,T \qquad (1)$$

where Q_t is the supply of wheat from the farmers of a particular country, P_t is the price of wheat (£) and Z_t is a measure of rainfall in this country. These three variables are observed. Price expectations P_t^* are revised by $P_t^*=\lambda P_{t-1}+(1-\lambda)P_{t-1}^*$. u_t is a random error, such that $E(u_t)=0$; $E(u_t^2)=\sigma^2$ and $E(u_su_t)=0$ if $s\neq t$ for all $s,t=1,2,....,T$.

(a) Show that this model can be reduced to infinite distributed lags model of the type

$$Q_t=\alpha+\beta\sum_{j=0}^{\infty}\mu^jP_{t-j-1}+\gamma Z_t+\varepsilon_t \qquad (2)$$

Solution

Starting from $P_t^*=\lambda P_{t-1}+(1-\lambda)P_{t-1}^*$

one can substitute lagged expression $P_{t-1}^*=\lambda P_{t-2}+(1-\lambda)P_{t-2}^*$ into it and get

$P_t^*=\lambda P_{t-1}+(1-\lambda)\lambda P_{t-2}+(1-\lambda)^2P_{t-2}^*$, then repeating this procedure

$P_t^*=\lambda P_{t-1}+(1-\lambda)\lambda P_{t-2}+(1-\lambda)^2\lambda P_{t-3}+(1-\lambda)^3P_{t-3}^*$ and so on until

$$P_t^* = \lambda P_{t-1} + (1-\lambda)\lambda P_{t-2} + (1-\lambda)^2 \lambda P_{t-3} + ... + (1-\lambda)^{k-1}\lambda P_{t-k} + (1-\lambda)^k P_{t-k}^*$$

When $k \to \infty$ supposing $(1-\lambda) < 1$ we get $(1-\lambda)^k P_{t-k}^* \to 0$ and

$$Q_t = \alpha + \beta(\lambda P_{t-1} + (1-\lambda)\lambda P_{t-2} + (1-\lambda)^2 \lambda P_{t-3} + ... + (1-\lambda)^{k-1}\lambda P_{t-k} + ...) + \gamma Z_t + u_t$$

Denoting $\mu^j = (1-\lambda)^j \lambda$ we get $Q_t = \alpha + \beta \sum_{j=0}^{\infty} \mu^j P_{t-j-1} + \gamma Z_t + u_t$

(b) How the models of the type (2) could be estimated? Discuss the problems connected with estimation of obtained model.

Solution

Disturbance term of this model inherits all properties of the original model.

The problem of estimating is the presence of the infinite lags while any data is finite, so we should stop substitution at certain point and omit unobservable term $(1-\lambda)^k P_{t-k}^*$ that is supposed to be small enough

$$Q_t = \alpha + \beta(\lambda P_{t-1} + (1-\lambda)\lambda P_{t-2} + (1-\lambda)^2 \lambda P_{t-3} + ... + (1-\lambda)^{k-1}\lambda P_{t-k}) + \gamma Z_t + u_t$$

The direct estimation of this model using OLS is impossible because of non-linearity in parameters and the conflicting estimates of coefficients, but it is possible to use non-linear least squares method. As P_{t-m} for different m are correlated the main problem in this estimation is multicollinearity. Another problem is the loss of k degrees of freedom (the loss of observations due to the using of lagged variables) that makes estimators inefficient.

Alternative approach is grid search: guessing the value of λ from 0 to 1 one can evaluate the sums $A_t(\lambda) = \lambda P_{t-1} + (1-\lambda)\lambda P_{t-2} + (1-\lambda)^2 \lambda P_{t-3} + ... + (1-\lambda)^{k-1}\lambda P_{t-k}$ and then use obtained values to estimate equation $Q_t = \alpha + \beta A_t(\lambda) + \gamma Z_t + u_t$ and to find $RSS(\lambda)$ of the equation; making this several times for different values of λ we then choose equation with the minimum $RSS(\lambda)$. This eliminates the problem of multicollinearity but the problem of reduced degrees of freedom and so inefficiency of estimators still remains.

(c) Show that the infinite distributed lags model of the type (2) can be written in terms of a single lag P_{t-1}. What estimation problems may occur in this model?

Solution

Starting from $Q_t = \alpha + \beta\lambda P_{t-1} + \beta(1-\lambda)\lambda P_{t-2} + \beta(1-\lambda)^2 \lambda P_{t-3} + ... + \beta(1-\lambda)^{k-1}\lambda P_{t-k} + ...) + \gamma Z_t + u_t$ one can lag it and then multiply by $(1-\lambda)$

$$(1-\lambda)Q_{t-1} = \alpha(1-\lambda) + \beta(1-\lambda)\lambda P_{t-2} + \beta(1-\lambda)^2 \lambda P_{t-3} + ... + \beta(1-\lambda)^{k-1}\lambda P_{t-k} + ...) + \gamma(1-\lambda)Z_{t-1} + (1-\lambda)u_{t-1}$$

and subtract the second equation from the first one; we get

$Q_t - (1-\lambda)Q_{t-1} = \alpha - \alpha(1-\lambda) + \beta\lambda P_{t-1} + \gamma(\lambda Z_t - (1-\lambda)Z_{t-1}) + u_t - (1-\lambda)u_{t-1}$ or

$$Q_t = \alpha\lambda + \beta\lambda P_{t-1} + \gamma(Z_t - (1-\lambda)Z_{t-1}) + (1-\lambda)Q_{t-1} + u_t - (1-\lambda)u_{t-1}$$

This is ADL(1,1) type model and usually model of this type could be estimated using OLS. But there are still some conflicting coefficients ($\gamma Z_t, \gamma(1-\lambda)Z_{t-1}$ and $(1-\lambda)Q_{t-1}$ terms) so NLS required.

The problem with this estimation is also the fact that *RHS* variable Q_{t-1} which is a function of u_{t-1} is correlated with the error term so *OLS* produces inconsistent parameter estimates.

(d) Derive an econometric model in observable quantities directly using Koyck transformation (without using infinite lags). Discuss the problems connected with estimation of obtained model.

Solution

From $P_t^* = \lambda P_{t-1} + (1-\lambda)P_{t-1}^*$

$P_t^* - (1-\lambda)P_{t-1}^* = \lambda P_{t-1}$, hence from

$Q_t = \alpha + \beta P_t^* + \gamma Z_t + u_t$ and lagged form

$Q_{t-1} = \alpha + \beta P_{t-1}^* + \gamma Z_{t-1} + u_{t-1}$ multipliyng the lagged form by $(1-\lambda)$ we get

$$Q_t - (1-\lambda)Q_{t-1} = [\alpha - (1-\lambda)\alpha] + \beta(P_t^* - (1-\lambda)P_{t-1}^*) + \gamma(Z_t - (1-\lambda)Z_{t-1}) + u_t - (1-\lambda)u_{t-1} =$$

$$= \alpha\lambda + \beta\lambda P_{t-1} + \gamma Z_t - (1-\lambda)Z_{t-1} + u_t - (1-\lambda)u_{t-1}.$$

$$Q_t = \alpha\lambda + (1-\lambda)Q_{t-1} + \beta\lambda P_{t-1} + \gamma Z_t - (1-\lambda)Z_{t-1} + u_t - (1-\lambda)u_{t-1}.$$

Note that the error term $\varepsilon_t = u_t - (1-\lambda)u_{t-1}$ is serially correlated (negative autocorrelation of the moving average type). Indeed

$\varepsilon_t = u_t - (1-\lambda)u_{t-1}$ while $\varepsilon_{t-q1} = u_{t-1} - (1-\lambda)u_{t-2}$, so under assumptions of the original model

$$\operatorname{cov}(\varepsilon_t;\varepsilon_{t-1}) = \operatorname{cov}(u_t - (1-\lambda)u_{t-1}; u_{t-1} - (1-\lambda)u_{t-2}) = \operatorname{cov}(-(1-\lambda)u_{t-1}; u_{t-1}) = -(1-\lambda)\operatorname{cov}(u_{t-1};u_{t-1}) = -(1-\lambda)\sigma_u^2$$

Therefore OLS estimators of coefficients will be inconsistent.

Question 7.

A researcher has time series data for aggregate consumption, *C*, and aggregate disposable personal income, *Y*, for a certain country. She establishes that the logarithms of both series are I(1) (integrated of order one). The researcher also believes that $\log C_t$ and $\log Y_t$ are cointegrated, so the cointegrating relationship exists

$$\log C_t = \alpha_1 + \alpha_2 \log Y_t + u_t \qquad (1)$$

where it may be assumed that u_t is normally distributed with zero mean and constant variance.

(a) Explain what is meant when a series is described as being integrated of order one. Explain what is meant when two time series are described as being cointegrated in the context of the problem under consideration. How to test both statements (the series is integrated of order 1, and two series under consideration are cointegrated)? Show that while $\log C_t$ and $\log Y_t$ are not stationary the logarithmic growth rates of consumption $\log\frac{C_t}{C_{t-1}}$ and income $\log\frac{Y_t}{Y_{t-1}}$ are stationary. Can be a regression of one of them on another considered as cointegrating relationship?

Solution

a) The time series is I(1) (integrated of order one) if this time series itself is non-stationary but its difference is stationary. To test this augmented Dickey-Fuller test is applied first to levels of original time series and then to the difference. In both cases the null hypothesis is non-stationarity.

Two time series are cointegrated if there exists a linear combination of the two series that is stable in the sense that the divergences, represented by the disturbance term, are stationary. The researcher should test both time series, their residuals and the difference $\log C_t - \hat{\alpha}_1 - \hat{\alpha}_2 \log Y_t$ for stationarity using a standard unit root test where $\hat{\alpha}_1, \hat{\alpha}_2$ are OLS estimators of the coefficients of regression (1).

The regression (1)

$$\log C_t = \alpha_1 + \alpha_2 \log Y_t + u_t$$

after lagging

$$\log C_{t-1} = \alpha_1 + \alpha_2 \log Y_{t-1} + u_{t-1}$$

and subtracting

$$\log C_t - \log C_{t-1} = \alpha_2(\log Y_t - \log Y_{t-1}) + u_t - u_{t-1}$$

gives

$$\log\frac{C_t}{C_{t-1}} = \alpha_2 \log\frac{Y_t}{Y_{t-1}} + u_t - u_{t-1} \quad (1^*)$$

Both logarithmic growth rates $\log\frac{C_t}{C_{t=1}} = \log C_t - \log C_{t-1} = \Delta \log C_t$ and $\log\frac{Y_t}{Y_{t=1}} = \log Y_t - \log Y_{t-1} = \Delta \log Y_t$ are stationary by assumptions of the model.
The obtained relationship (1*) between logarithmic growth rates cannot be considered as cointegrating relationship as cointegration is not applicable to the stationary time series.

(b) The researcher is also interested in the short-run dynamics of the relationship and correctly hypothesizes that they may be represented by ADL(1,2) process with two lags for income variable

$$\log C_t = \beta_1 + \beta_2 \log C_{t-1} + \beta_3 \log Y_t + \beta_4 \log Y_{t-1} + \beta_5 \log Y_{t-2} + \varepsilon_t \qquad (2)$$

where ε_t is identically and independently distributed and drawn from a normal distribution with zero mean. Derive long run relationship between the equilibrium values of $\log C_t$ and $\log Y_t$: $\overline{\log C}$ and $\overline{\log Y}$, and find short run and long run income elasticities of consumption.

Solution

Putting $\log C_t = \log C_{t-1} = \overline{\log C}$ and $\log Y_t = \log Y_{t-1} = \log Y_{t-2} = \overline{\log Y}$, (2) implies the long-run relationship

$$\overline{\log C} = \beta_1 + \beta_2 \overline{\log C} + \beta_3 \overline{\log Y} + \beta_4 \overline{\log Y} = \frac{\beta_1}{1-\beta_2} + \frac{\beta_3+\beta_4+\beta_5}{1-\beta_2}\overline{\log Y}$$

So β_3 represents short run elasticity while long run elasticity is given by expression $\frac{\beta_3+\beta_4+\beta_5}{1-\beta_2}$. Here short run elasticity can be greater or less than long run elasticity depending on the values of the coefficients $\beta_2, \beta_3, \beta_4, \beta_5$.

(c) Using long run relationship between $\log C_{t-1}$ and $\log Y_{t-1}$ as a possible cointegrating relationship derive error-correction type model for the model (2).

Solution

Let us make a plan of transformations needed. From the previous result

$$\overline{\log C} = \frac{\beta_1}{1-\beta_2} + \frac{\beta_3+\beta_4+\beta_5}{1-\beta_2}\overline{\log Y}$$

So cointegrating relationship should have the form

$$\log C_{t-1} - \frac{\beta_1}{1-\beta_2} - \frac{\beta_3+\beta_4+\beta_5}{1-\beta_2}\log Y_{t-1}$$

From here the error correction style model should be like this

$$\Delta \log C_t = (\beta_2 - 1)\left(\log C_{t-1} - \frac{\beta_1}{1-\beta_2} - \frac{\beta_3+\beta_4+\beta_5}{1-\beta_2}\log Y_{t-1}\right) + some \;\; differences + \varepsilon_t$$

Expanding this equation we find

$$\log C_t = \beta_1 + \beta_2 \log C_{t-1} + \underline{\beta_3 \log Y_{t-1}} + \beta_4 \log Y_{t-1} + \underline{\beta_5 \log Y_{t-1}} + some \;\; differences + \varepsilon_t$$

Compare this expression with ADL(1,1) model

$$\log C_t = \beta_1 + \beta_2 \log C_{t-1} + \underline{\underline{\beta_3 \log Y_t}} + \beta_4 \log Y_{t-1} + \underline{\underline{\beta_5 \log Y_{t-2}}} + \varepsilon_t$$

Underlined are expressions that should not be in ADL model – they should be duplicated with the sign 'minus' in difference terms, double underlined expressions, which are lacking for the model – they should be duplicated with the sign 'plus' in difference terms,. Now we can find the differences in the right hand side of the model

$$\Delta \log C_t = (\beta_2 - 1)\left(\log C_{t-1} - \frac{\beta_1}{1-\beta_2} - \frac{\beta_3 + \beta_4 + \beta_5}{1-\beta_2} \log Y_{t-1} \right) +$$

$$+ \beta_3 \log Y_t - \beta_3 \log Y_{t-1} - \beta_5 \log Y_{t-1} + \beta_5 \log Y_{t-2} + \varepsilon_t$$

or finally

$$\Delta \log C_t = (\beta_2 - 1)\left(\log C_{t-1} - \frac{\beta_1}{1-\beta_2} - \frac{\beta_3 + \beta_4 + \beta_5}{1-\beta_2} \log Y_{t-1} \right) + \beta_3 \Delta \log Y_t - \beta_5 \Delta \log Y_{t-1} + \varepsilon_t$$

(d) Describe the structure and dynamics of error correction model. Explain why fitting the error-correction model, rather than (2) directly, avoids a potentially important problem. Which assumptions are necessary to use error correction model? How these assumptions could be tested?

Solution

Any error correction model includes two parts: basic model using first and possibly higher order differences

$$\Delta \log C_t = \beta_3 \Delta \log Y_t - \beta_5 \Delta \log Y_{t-1} + \varepsilon_t$$

and the error correction mechanism

$$(\beta_2 - 1)\left(\log C_{t-1} - \frac{\beta_1}{1-\beta_2} - \frac{\beta_3 + \beta_4 + \beta_5}{1-\beta_2} \log Y_{t-1} \right)$$

based on cointegration relationship. Error correction mechanism allows to compensate partially errors caused by the myopia of the short term relationship described by difference model. The speed of correction depends on the value of the coefficient $\beta_2 - 1$ in the error correction mechanism.

The advantage of fitting the error-correction model, rather than original model

$$\log C_t = \alpha_1 + \alpha_2 \log Y_t + u_t \quad (1)$$

directly, is that the variables in this relationship are all stationary and so the relationship may be fitted with standard OLS. Estimation and further analysis of original model based on non-stationary time series could lead to spurious regression.

The set of tests should be performed: the original time series should be non stationary, their first differences - stationary. The residuals of the regression $\log C_t = \alpha_1 + \alpha_2 \log Y_t + u_t$ should be also stationary).

The error correction model in this case uses first differences, and cointegrating relationship, so all parts of it are stationary.

The International College of Economics and Finance
Econometrics – 2016-2017.
Final exam 2017. May 24.

Suggested solutions

General instructions. **Candidates should answer SIX of the following SEVEN questions: all 4 questions of the Section A and any 2 of the questions from Section B (questions 5-7). The weight of the Section A is 60% of the exam; two other questions from the Section B add 20% each. You are advised to divide your time accordingly. Structure your answers in accordance with the structure of the questions. When testing hypotheses always state clearly null and alternative hypotheses providing critical value used for test, mentioning degrees of freedom and the significance level chosen for the test.**

SECTION A
(1 hour 40 minutes)

Answer **ALL** questions from this section (questions **1-4**).
Each question in this section bears 15 marks

Question 1.

The researcher is interested in estimation of influence of the tenure with the current employer in years (*TENURE*) on the earnings in dollars per hour (*EARN*). She starts with estimating simple linear regression on the base of the data on 50 respondents from the National Longitudinal Survey of Youth for the year 2000:

$$\underset{(1,55)}{EARN = 9.95} + \underset{(0.21)}{0.61TENURE} \qquad R^2 = 0.12 \tag{1}$$

Then she adds the years of schooling (*HGC*) as explanatory variable. She also suspect that the excessive weight of the respondent could be obstacle for achieving good earnings, so she includes additional explanatory variable *WEIGHT* (measured in pounds)

$$\underset{(6,78)}{EARN = -6.08} + \underset{(0.21)}{0.46TENURE} + \underset{(0.42)}{1.14HGC} + \underset{(0.02)}{0.01WEIGHT} \qquad R^2 = 0.23 \tag{2}$$

(a) Are the slope coefficients of the variable *TENURE* in regressions (1) and (2) significant? What do you mean by significance and what practical implications follow from the significance of the slope coefficient? Why the coefficients of *TENURE* are different in equations (1) and (2)? Is there any difference in interpretation of these coefficients?

Solution

t-statistics for the slope in equation (1) is $t = \frac{0.61}{0.21} = 2.9$. $2.9 > 2.68 = t(two\ sided, crit., 1\%, df = 48)$, so the coefficient is significant. Its significance means that it could be considered as non-zero (null hypothesis $H_0: \beta_2 = 0$ is rejected at 1% significance level). This implies that *TENURE* influence earnings, so any analytical work involving the value of this coefficient is reasonable.

t-statistics for the coefficient of *TENURE* in equation (2) is now $t = \frac{0.46}{0.21} = 2.19$, that is significant only at 5% significance level:

$2.69 = t(two\ sided, crit., 1\%, df = 46) > 2.19 > 2.01 = t(two\ sided, crit., 5\%, df = 46)$.

The difference of values in coefficients of *TENURE* in (1) and (2) is explained by the fact that in (2) some additional variables are present so the influence of *TENURE* could be measured more precisely. The coefficient in equation (2) shows that *EARN* rises by 46 cents per hour if tenure rises by 1 year holding other explanatory variables constant.

(b) Is the researcher right in her suggestion concerning variable *WEIGHT*? Explain, using appropriate test(s). Are two variables *HGC* and *WEIGHT* taken together significant?

Solution

While the coefficient of *HGC* is significant at 1% level: $t = \frac{1.14}{0.42} = 2.71 > 2.69$, the coefficient of *WEIGHT* is insignificant. It should be tested using one sided critical value but as it is in fact positive the null hypothesis non-positive influence of *WEIGHT* is not rejected in any case.

To test joint significance of two variables *HGC* and *WEIGHT* one should use F-test that can be evaluated as $F = \frac{(R_U^2 - R_R^2)/(number\ of\ added\ variables)}{(1-R_U^2)/df} = \frac{(0.23-0.12)/2}{(1-0.23)/46} = 3.28$, what is greater than $3.23 = F(crit., 5\%, df1 = 2, df2 = 40)$ so *HGC* and *WEIGHT* are jointly significant.

(c) What equation (1) or (2) seems to be more reliable and adequate? What are relative advantages and disadvantages of choosing equations (1) or (2) as correct specification? Discuss both options in turn. Based on the previous discussion what equation would you propose for the further analysis and why?

Solution

Joint significance of two variables *HGC* and *WEIGHT* in (b) indicates that equation (2) is more relevant to the investigated problem. Nevertheless both equations suffer from several flaws so the situation is rather ambiguous. If equation (2) is assumed to be correctly specified the estimates of equation (1) could be biased and standard errors and test could be invalid. In this case for example the value of slope in equation (1) equal to 0.61 could be considered as overestimated (in equation (2) the corresponding coefficient is only 0.46).

In turn if specification (1) is correct one can interpret the insignificance of *WEIGHT* as possible sign of multicollinearity.

Taking into account all results and considerations one can suggest alternative approach based on the assumption that the variable *WEIGHT* is not relevant at all to discussed problem. So one could advice the researcher to estimate equation $EARN = \beta_1 + \beta_2 TENURE + \beta_3 HGC + u_i$. One could hope that this specification would combine advantages of both specifications (1) and (2).

Question 2.

A student has data from the National Longitudinal Survey of Youth for the year 2000 on full years of schooling, *HGC*, and composite test of intellectual abilities, *ASVABC*, for a sample of 22 males and 15 females. She defines the variable *EIA* (excessive intellectual abilities = excess over minimum value of *ASVABC*) as *ASVABC* – 30 as she noticed that minimum value of *ASVABC* in the sample was 30). She defines also dummy variable *MALE* for being male, a slope dummy variable *MALEEIA* as the product of *MALE* and *EIA*. She performs the regressions (1) (2) and (3) for the entire sample (standard errors in parentheses):

$$\underset{}{HGC} = \underset{(0,71)}{9.68} + \underset{(0.03)}{0.16}EIA \qquad R^2 = 0.44 \tag{1}$$

$$HGC = \underset{(0,76)}{10.20} + \underset{(0.03)}{0.16}EIA - \underset{(0.62)}{1.00}MALE \qquad R^2 = 0.48 \tag{2}$$

$$HGC = \underset{(1,07)}{8.63} + \underset{(0.05)}{0.24}EIA + \underset{(1.37)}{1.44}MALE - \underset{(0.06)}{0.12}MALEEIA \qquad R^2 = 0.54 \tag{3}$$

(a) Give an interpretation of all coefficients in regression (1) in the context of the problem of gender differences. Give the interpretation of all coefficients in regression (2) in the same context. Having in mind the same context give the interpretation of all coefficients in regression (3).

Solution

9.68 in equation **(1)** is the value of schooling (in years on average) for a person (no matter male or female she or he is) that is characterized by minimal value of ASVABC = 30 points. Each additional point in ASVABC exceeding 30 adds to this value 0.16 years.

Including dummies allows to analyze additionally the influence of gender. Equation **(2)** allows only to get different estimates for intercept for men and women under assumption that marginal effect of EIA (and so ASVABC) on HGC is the same for men and women (estimated as 0.16). As women are reference category here estimated intercept 10.2 years refers to women while for men it is less by 1 year (but this difference has turned to be insignificant). This insignificance is possibly caused by invalid assumption of equal marginal effects not depending of gender.

Equation **(3)** allows to take into account additionally the difference in marginal effects. Now equation for females (reference category) is $HGC = 8.63 + 0.24EIA$. Its interpretation is similar to that of equation (1) but relates only to females. The equation $HGC = (8.63 + 1.44) + (0.24 - 0.12)EIA$ relates to males.

(b) Is there a significant difference between males and females in the overall influence of intellectual abilities on the years of schooling? Perform appropriate tests and discuss the results on the basis of comparison of equations (1), (2) and (3).

Solution

As it was mentioned in (a) the insignificance of *MALE* in equation (2) is possibly caused by invalid assumption of equal marginal effects not depending of gender.

To evaluate significance of all variables connected with gender in equation (2) we should perform F-test comparing two equations (1) and (2): $F = \dfrac{(0.54 - 0.44)/2}{(1 - 0.54)/(37 - 4)} = 3.58$ while the critical values are $F(1\%; 2; 35) = 5.27$ and $F(5\%; 2; 30) = 3.32$ so the null hypothesis is rejected at 5% significance level. One should be mentioned that both coefficients associated with gender are insignificant, while total influence of gender is significant.

(c) Alternative approach to detection of gender differences in the analysis of influence of *EIA* on *HGC* is the Chow test. Describe this test in details. What additional information is needed to perform this test? Is this test equivalent to any of the tests performed in **(b)**?

Solution

To conduct Chow test one should run regression (1) three times: firsts – for the whole sample memorizing the value of RSS_{tot}, second – for subsample of 22 men getting RSS_m and third – for subsample of 15 women getting RSS_w. Now it is possible to evaluate F-statistic $F = \dfrac{(RSS_{tot} - (RSS_m + RSS_w))/2}{(RSS_m + RSS_w)/33}$.

Chow test is equivalent to the F-test performed in **(b)** to compare equations (1) and (3) (with full set of dummies).

Question 3.

(a) What is meant by a stationary time series? What is meant by a non-stationary time series? Give two examples of stationary time series and two examples of non-stationary time series.

Solution

A time series is stationary (weakly) if the mean, variance and autocovariance of the series are independent of time. Autocovariance may depend on length of the lag. The violation of any of these assumptions makes the series non-stationary.

Two examples of stationary time series are autoregression $X_t = \theta X_{t-1} + u_t \quad |\theta| < 1$, and moving average $X_t = u_t + \mu u_{t-1}$.

Two examples of non-stationary time series are random walk $X_t = X_{t-1} + u_t$, and time trend $X_t = \alpha_1 + \alpha_2 t + u_t$.

All examples above use the same assumptions on u_t: $E(u_t) = 0$; $E(u_t^2) = \sigma^2$ and $E(u_s u_t) = 0$ for $s \neq t$.

(b-c) Under what conditions is the series generated by $X_t = \theta X_{t-1} + u_t$ stationary? Explain your answer, evaluating expected value and variance **(b)** and then covariance **(c)**. Assume X_0 is fixed. $E(u_t) = 0$; $\mathrm{Var}(u_t) = \sigma^2$ and $E(u_s u_t) = 0$ for all *s* and *t*, $s \neq t$.

Solution

(b) We can write: $t = 1$: $X_1 = \theta X_0 + u_1$,

$$t = 2:\ X_2 = \theta X_1 + u_2 = \theta(\theta X_0 + u_1) + u_2 = \theta^2 X_0 + \theta u_1 + u_2,$$

$$t = 3:\ X_3 = \theta X_2 + u_3 = \theta^3 X_0 + \theta^2 u_1 + \theta u_2 + u_3 \qquad ,$$

$$\vdots$$

Doing these recursive substitutions, we can write:

$$X_t = \theta X_{t-1} + u_t = \theta^t X_0 + u_t + \theta u_{t-1} + \ldots + \theta^{t-1} u_1 .$$

Therefore:

$$E(X_t) = E(\theta^t X_0 + u_t + \theta u_{t-1} + \ldots + \theta^{t-1} u_1) = \theta^t X_0,$$

and $\mathrm{var}(X_t) = \mathrm{var}(\theta^t X_0 + u_t + \theta u_{t-1} + \ldots + \theta^{t-1} u_1) = \sigma_\varepsilon^2 (1 + \theta^2 + \theta^4 + \ldots + \theta^{2(t-1)}) = \sigma^2 \sum_{s=0}^{t-1} \theta^{2s}$.

If $\theta \geq 1$ then for large '*t*', it is easy to see that $E(X_t) \to \infty$ and $\mathrm{Var}(X_t) \to \infty$. So the variable X_t is non-stationary and standard test statistics are not valid as using these statistics assume stationarity and finite mean and variance for the random variable. If this is the case then the variable is growing at the exponential rate which is rare for economic variables.

If $|\theta| < 1$, then for large '*t*': $E(X_t) = \theta^t X_0 = 0$,

and $\mathrm{var}(X_t) = \sigma^2 \sum_{s=0}^{t-1} \theta^{2s} = \sigma^2 (1 + \theta^2 + \theta^4 + \ldots) = \dfrac{\sigma^2}{1 - \theta^2}$, which is a constant.

(c) For large *t*, covariance is given by:

$$\mathrm{cov}(X_t, X_{t-s}) = E[X_t - E(X_t)][X_{t-s} - E(X_{t-s})] = E[X_t X_{t-s}] =$$

$$= E[u_t + \theta u_{t-1} + \theta^2 u_{t-2} + \ldots + \theta^{t-1} u_1][u_{t-s} + \theta u_{t-s-1} + \ldots + \theta^{t-1} u_{-s}] =$$

$$= \theta^s (1 + \theta^2 + \theta^4 + \ldots)\sigma^2 = \frac{\theta^s \sigma^2}{1 - \theta^2},$$

which depends only upon the value of 's'. Therefore the variable X_t is stationary if $|\theta| < 1$.

Question 4.

The student is interested in the influence of the price level and income on the consumption of strong drinks SD_{it}. Suppose that the time series V_t (vodka), C_t (cognac), W_t (whiskey), G_t (gin), B_t (brandy), R_t (rum) for 1987-2016 form the panel (as observations for particular units of some general type of strong drinks (SD_{it})). DPI_t is the disposable personal income and $PRSD_{it}$ is the relative price index for corresponding SD_{it} .Let the model under investigation be

$$\log(SD_{it}) = \beta_1 + \beta_2 \log(DPI_t) + \beta_3 \log(PRSD_{it}) + u_{it}.$$

(a) The student decided to use the approach based on the fixed effect panel data model. Help the student to understand what is LSDV method and how to use it for estimation of panel data model under consideration. Explain clearly what are fixed effects.

Solution

The main question in the panel data analysis is the problem of the origin of unobserved heterogeneity $\sum_{p=1}^{s}\gamma_p Z_{pi}$ = α_i in the model of the type

$$Y_{it} = \beta_0 + \sum_{j=1}^{k}\beta_j X_{jit} + \sum_{p=1}^{s}\gamma_p Z_{pi} + \varepsilon_{it} \qquad (1).$$

Fixed effect approach assigns it to the fixed characteristics of individual elements (here to the types of strong drinks), while random effect approach assigns it to the random factors.

In the **Least Squares Dummy Variable (LSDV) fixed effect method** the unobserved effect is brought explicitly into the model. A set of dummy variables D_i is defined, where D_i is equal to 1 in the case of an observation relating to an individual *i* and 0 otherwise. The model can be written as

$$Y_{it} = \sum_{j=2}^{K}\beta_j X_{jit} + \sum_{i=1}^{n}\alpha_i D_i + u_{it}. \qquad \text{(vi)}$$

The unobserved effect is now being treated as the coefficient of the specific individual *i*.

The term $\alpha_i D_i$ represents a fixed effect on the dependent variable Y_i for individual *i*. We use the model without intercept otherwise we will fall into the dummy variable trap. Entering the details, we note that the method calculates its own intercept as an average of all coefficients for dummies, and represents fixed effects as the deviations of these coefficients from the calculated mean. The method is based on the assumption that the elasticities in income and prices are equal for all strong drinks, and heterogeneity is reflected only in the intercept.

It should be noted that inclusion of a set of dummies leads to losing of additional degrees of freedom, so estimators become less efficient.

(b) The model above is based on the assumption that the elasticities of different goods are the same. Suggest how to test this assumption. Give some details: indicate corresponding test and the data needed for it, null hypothesis, distribution of the test statistic, the number of degrees of freedom and the decision rule.

Solution

We start from LSDV fixed effects model of the type

$$\log(SD_{it}) = \beta_2 \log(DPI_t) + \beta_3 \log(PRSD_{it}) + \sum_{i=1}^{6}\gamma_i D_i + u_{it}. \qquad t = 1, 2, ..., 30 \quad (1)$$

where D_i are dummies corresponding different drinks under consideration. To take into account that the elasticities of DPI_t and $PRSD_{it}$ could be different, it is possible to run instead 6 different regressions for 6 drinks ($i = 1, 2, ..., 6$),

$$\log(SD_{it}) = \gamma_1 + \gamma_2 \log(DPI_t) + \gamma_3 \log(PRSD_{it}) + u_{it}$$

and then perform an F-test using the values of RSS' of evaluated models:

$$F=\frac{(RSS_{LSDV}-\sum_{i=1}^{6}RSS_i)/(6\cdot 3-(2+6))}{\sum_{i=1}^{6}RSS_i/(30-6\cdot 3)}=\frac{(RSS_{LSDV}-\sum_{i=1}^{6}RSS_i)/10}{\sum_{i=1}^{6}RSS_i/12}.$$

Here RSS_{LSDV} - the value of RSS for LSDV fixed effect model, $\sum_{i=1}^{6}RSS_i$ - sum of RSS' for separate regressions for different goods, the rule for degrees of freedom follows general principles of F-test and is clear from the formula above. If H_0 is not rejected there is no significant differences between elasticities.

(c) The alternative approach is random effect panel data model. How to decide what approach (fixed effect or random effect) is better? Indicate corresponding test, null hypothesis, distribution of the test statistic, the number of degrees of freedom and the decision rule (which alternative is chosen if null hypothesis is rejected and which if it is not). What are advantages and risks of each choice?

Solution

The student is recommended to use Darbin-Wu-Hausman (DWH) test to choose between fixed and random effects. It is standard for majority of econometric computer programs and is based on using chi-square statistics with degrees of freedom equal to the number of variables in the equation under consideration (2 in our case) (as it compares estimates of coefficients obtained by two alternative models). Under H_0 that there is no difference between coefficients obtained by two alternative models – fixed and random panel models (which means that unobserved heterogeneity α_i as a part of disturbance term, is not correlated with DPI_t and $PRSD_{it}$) both fixed effect and random effect models provide us with consistent estimates. We choose in this case random effect models as it retains in disturbance term all unobserved heterogeneity, there is no reduction of degrees of freedom typical for fixed effects models.

If H_0 is rejected, so there are essential differences between coefficients obtained using fixed and random effects models, we choose fixed effects model, because rejecting of H_0 means that main assumption of independence of the disturbance term from regressors is violated so using random effects model we are under risk of getting inconsistent estimates of parameters. So we have to suffice the fixed effects model that always gives consistent estimates.

SECTION B

(1 hour)

Answer **TWO** questions from this section (questions **5-7**).

Each answered question in this section bears 20 marks.

Question 5.

Suppose the model $Y_i = \beta_1 + \beta_2 X_i + u_i,\ \ i = 1, \ldots, n$ satisfies all assumptions of the model A (Gauss-Markov conditions). Let the values of X_i are supposed to be non-stochastic. The researcher wrongly believes that $\beta_1 = 0$ and so uses OLS estimator $\tilde{\beta}_2$ for the model $Y_i = \beta_2 X_i + u_i,\ \ i = 1, \ldots, n$ instead of OLS estimator $\hat{\beta}_2 = \dfrac{\sum x_i y_i}{\sum x_i^2}$ for the model $Y_i = \beta_1 + \beta_2 X_i + u_i,\ \ i = 1, \ldots, n$ where $x_i = X_i - \overline{X}$, $y_i = Y_i - \overline{Y}$.

(a) Show that OLS estimator $\tilde{\beta}_2$ for the model $Y_i = \beta_2 X_i + u_i,\ \ i = 1, \ldots, n$ is $\tilde{\beta}_2 = \dfrac{\sum X_i Y_i}{\sum X_i^2}$.

Solution

Under assumption $\beta_1 = 0$ OLS is minimization of a function $F(\beta_2) = \sum (Y_i - \beta_2 X_i)^2$. First order condition for this is $\dfrac{d}{d\beta_2} \sum (Y_i - \beta_2 X_i)^2 = 0$ or $\sum (Y_i - \beta_2 X_i) X_i = 0$ that gives $\tilde{\beta}_2 = \dfrac{\sum X_i Y_i}{\sum X_i^2}$

(b) If the assumption $\beta_1 = 0$ is not correct show that estimator $\tilde{\beta}_2 = \dfrac{\sum X_i Y_i}{\sum X_i^2}$ is biased and find the bias. Indicate the case when despite $\beta_1 \neq 0$ the estimator $\tilde{\beta}_2 = \dfrac{\sum X_i Y_i}{\sum X_i^2}$ is nevertheless unbiased.

Solution

$E\tilde{\beta}_2 = E\dfrac{\sum X_i Y_i}{\sum X_i^2} = \dfrac{\sum X_i EY_i}{\sum X_i^2} = \dfrac{\sum X_i E(\beta_1 + \beta_2 X_i + u_i)}{\sum X_i^2} = \dfrac{\sum X_i (\beta_1 + \beta_2 X_i + Eu_i)}{\sum X_i^2}$. According to G.M.C. $Eu_i = 0$ so $E\tilde{\beta}_2 = \dfrac{\sum X_i (\beta_1 + \beta_2 X_i)}{\sum X_i^2} = \dfrac{\sum \beta_1 X_i + \sum \beta_2 X_i^2}{\sum X_i^2} = \beta_2 + \beta_1 \dfrac{\sum X_i}{\sum X_i^2}$. The bias is $\beta_1 \dfrac{\sum X_i}{\sum X_i^2}$.

The bias disappears if $\sum X_i = 0$ under assumption $\beta_1 \neq 0$.

(c) Show that the variance of $\tilde{\beta}_2 = \dfrac{\sum X_i Y_i}{\sum X_i^2}$ is given by expression $\operatorname{var}(\tilde{\beta}_2) = \dfrac{\sigma^2}{\sum X_i^2}$, where $\sigma^2 = \operatorname{var}(u_i)$.

Solution

$$\operatorname{var}(\tilde{\beta}_2) = \operatorname{var}\left(\frac{\sum_{t=1}^{T} X_t Y_t}{\sum_{t=1}^{T} X_t^2}\right) = \frac{\operatorname{var}\left(\sum_{t=1}^{T} X_t Y_t\right)}{\left(\sum_{t=1}^{T} X_t^2\right)^2} = \frac{\sum_{t=1}^{T} (\operatorname{var} X_t Y_t)}{\left(\sum_{t=1}^{T} X_t^2\right)^2} = \frac{\sum_{t=1}^{T} X_t^2 (\operatorname{var} Y_t)}{\left(\sum_{t=1}^{T} X_t^2\right)^2} = \frac{\sum_{t=1}^{T} X_t^2 \sigma^2}{\left(\sum_{t=1}^{T} X_t^2\right)^2} = \frac{\sigma^2}{\sum_{t=1}^{T} X_t^2}$$

(d) Show that if assumption $\beta_1 = 0$ is true the estimator $\hat{\beta}_2 = \dfrac{\sum x_i y_i}{\sum x_i^2}$ is still unbiased but inefficient.

(Hint: compare the variances of two estimators $\tilde{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2}$ and $\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$).

Solution

As it is known $\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} = \beta_2 + \frac{\sum x_i u_i}{\sum x_i^2}$. Under assumption $\beta_1 = 0$ $E\hat{\beta}_2 = \beta_2 + \frac{\sum x_i Eu_i}{\sum x_i^2} = \beta_2$, as $Eu_i = 0$, so the estimator still remains unbiased.

It is known that $\mathrm{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$ and we proved in **(c)** that $\mathrm{var}(\tilde{\beta}_2) = \frac{\sigma^2}{\sum X_i^2}$ ..So we need to compare both variances. Note that $\sum x_i^2 = \sum (X_i - \overline{X})^2 = \sum X_i^2 - 2\overline{X} \sum X_i + \sum (\overline{X})^2 = \sum X_i^2 - n(\overline{X})^2$ and obviously $\sum X_i^2 - n\sum (\overline{X})^2 \le \sum X_i^2$ so $\mathrm{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} = \frac{\sigma^2}{\sum X_i^2 - n(\overline{X})^2} \ge \frac{\sigma^2}{\sum X_i^2} = \mathrm{var}(\tilde{\beta}_2)$.

That means that if assumption $\beta_1 = 0$ is correct $\hat{\beta}_2$ has bigger variance and so is no more efficient estimator of a slope.

Question 6.

Consider a model

$$Q_t = \alpha + \beta P_t^* + \gamma Z_t + u_t ;\ t = 1, 2, ..., T \qquad (1)$$

where Q_t is the supply of wheat from the farmers of a particular country, P_t is the price of wheat (£) and Z_t is a measure of rainfall in this country. These three variables are observed. Price expectations P_t^* are revised by $P_t^* = \lambda P_{t-1} + (1-\lambda) P_{t-1}^*$. u_t is a random error, such that $E(u_t) = 0$; $E(u_t^2) = \sigma^2$ and $E(u_s u_t) = 0$ if $s \neq t$ for all $s, t = 1, 2, ..., T$.

(a) Show that this model can be reduced to infinite distributed lags model of the type

$$Q_t = \alpha + \beta \sum_{j=0}^{\infty} \mu^j P_{t-j-1} + \gamma Z_t + \varepsilon_t \qquad (2)$$

Solution

Starting from $P_t^* = \lambda P_{t-1} + (1-\lambda) P_{t-1}^*$

one can substitute lagged expression $P_{t-1}^* = \lambda P_{t-2} + (1-\lambda) P_{t-2}^*$ into it and get

$P_t^* = \lambda P_{t-1} + (1-\lambda)\lambda P_{t-2} + (1-\lambda)^2 P_{t-2}^*$, then repeating this procedure

$P_t^* = \lambda P_{t-1} + (1-\lambda)\lambda P_{t-2} + (1-\lambda)^2 \lambda P_{t-3} + (1-\lambda)^3 P_{t-3}^*$ and so on until

$P_t^* = \lambda P_{t-1} + (1-\lambda)\lambda P_{t-2} + (1-\lambda)^2 \lambda P_{t-3} + ... + (1-\lambda)^{k-1} \lambda P_{t-k} + (1-\lambda)^k P_{t-k}^*$

When $k \to \infty$ supposing $(1-\lambda) < 1$ we get $(1-\lambda)^k P_{t-k}^* \to 0$ and

$Q_t = \alpha + \beta(\lambda P_{t-1} + (1-\lambda)\lambda P_{t-2} + (1-\lambda)^2 \lambda P_{t-3} + ... + (1-\lambda)^{k-1} \lambda P_{t-k} + ...) + \gamma Z_t + u_t$

Denoting $\mu^j = (1-\lambda)^j \lambda$ we get $Q_t = \alpha + \beta \sum_{j=0}^{\infty} \mu^j P_{t-j-1} + \gamma Z_t + u_t$

(b) How the models of the type (2) could be estimated? Discuss the problems connected with estimation of obtained model.

Solution

Disturbance term of this model inherits all properties of the original model.

The problem of estimating is the presence of the infinite lags while any data is finite, so we should stop substitution at certain point and omit unobservable term $(1-\lambda)^k P^*_{t-k}$ that is supposed to be small enough

$$Q_t = \alpha + \beta(\lambda P_{t-1} + (1-\lambda)\lambda P_{t-2} + (1-\lambda)^2 \lambda P_{t-3} + ... + (1-\lambda)^{k-1}\lambda P_{t-k}) + \gamma Z_t + u_t$$

The direct estimation of this model using OLS is impossible because of non-linearity in parameters and the conflicting estimates of coefficients, but it is possible to use non-linear least squares method. As P_{t-m} for different m are correlated the main problem in this estimation is multicollinearity. Another problem is the loss of k degrees of freedom (the loss of observations due to the using of lagged variables) that makes estimators inefficient.

Alternative approach is grid search: guessing the value of λ from 0 to 1 one can evaluate the sums $A_t(\lambda) = \lambda P_{t-1} + (1-\lambda)\lambda P_{t-2} + (1-\lambda)^2 \lambda P_{t-3} + ... + (1-\lambda)^{k-1}\lambda P_{t-k}$ and then use obtained values to estimate equation $Q_t = \alpha + \beta A_t(\lambda) + \gamma Z_t + u_t$ and to find $RSS(\lambda)$ of the equation; making this several times for different values of λ we then choose equation with the minimum $RSS(\lambda)$. This eliminates the problem of multicollinearity but the problem of reduced degrees of freedom and so inefficiency of estimators still remains.

(c) Show that the infinite distributed lags model of the type (2) can be written in terms of a single lag P_{t-1}. What estimation problems may occur in this model?

Solution

Starting from $Q_t = \alpha + \beta\lambda P_{t-1} + \beta(1-\lambda)\lambda P_{t-2} + \beta(1-\lambda)^2 \lambda P_{t-3} + ... + \beta(1-\lambda)^{k-1}\lambda P_{t-k} + ...) + \gamma Z_t + u_t$ one can lag it and then multiply by $(1-\lambda)$

$$(1-\lambda)Q_{t-1} = \alpha(1-\lambda) + \beta(1-\lambda)\lambda P_{t-2} + \beta(1-\lambda)^2 \lambda P_{t-3} + ... + \beta(1-\lambda)^{k-1}\lambda P_{t-k} + ...) + \gamma(1-\lambda)Z_{t-1} + (1-\lambda)u_{t-1}$$

and subtract the second equation from the first one; we get

$Q_t - (1-\lambda)Q_{t-1} = \alpha - \alpha(1-\lambda) + \beta\lambda P_{t-1} + \gamma(\lambda Z_t - (1-\lambda)Z_{t-1}) + u_t - (1-\lambda)u_{t-1}$ or

$$Q_t = \alpha\lambda + \beta\lambda P_{t-1} + \gamma(Z_t - (1-\lambda)Z_{t-1}) + (1-\lambda)Q_{t-1} + u_t - (1-\lambda)u_{t-1}$$

This is ADL(1,1) type model and usually model of this type could be estimated using OLS. But there are still some conflicting coefficients (γZ_t, $\gamma(1-\lambda)Z_{t-1}$ and $(1-\lambda)Q_{t-1}$ terms) so NLS required.

The problem with this estimation is also the fact that *RHS* variable Q_{t-1} which is a function of u_{t-1} is correlated with the error term so *OLS* produces inconsistent parameter estimates.

(d) Derive an econometric model in observable quantities directly using Koyck transformation (without using infinite lags). Discuss the problems connected with estimation of obtained model.

Solution

From $P^*_t = \lambda P_{t-1} + (1-\lambda)P^*_{t-1}$

$P^*_t - (1-\lambda)P^*_{t-1} = \lambda P_{t-1}$, hence from

$Q_t = \alpha + \beta P^*_t + \gamma Z_t + u_t$ and lagged form

$Q_{t-1} = \alpha + \beta P^*_{t-1} + \gamma Z_{t-1} + u_{t-1}$ multipliyng the lagged form by $(1-\lambda)$ we get

$$Q_t - (1-\lambda)Q_{t-1} = [\alpha - (1-\lambda)\alpha] + \beta(P^*_t - (1-\lambda)P^*_{t-1}) + \gamma(Z_t - (1-\lambda)Z_{t-1}) + u_t - (1-\lambda)u_{t-1} =$$
$$= \alpha\lambda + \beta\lambda P_{t-1} + \gamma Z_t - (1-\lambda)Z_{t-1} + u_t - (1-\lambda)u_{t-1}.$$

$$Q_t = \alpha\lambda + (1-\lambda)Q_{t-1} + \beta\lambda P_{t-1} + \gamma Z_t - (1-\lambda)Z_{t-1} + u_t - (1-\lambda)u_{t-1}.$$

Note that the error term $\varepsilon_t = u_t - (1-\lambda)u_{t-1}$ is serially correlated (negative autocorrelation of the moving average type). Indeed
$\varepsilon_t = u_t - (1-\lambda)u_{t-1}$ while $\varepsilon_{t-q1} = u_{t-1} - (1-\lambda)u_{t-2}$, so under assumptions of the original model
$\operatorname{cov}(\varepsilon_t;\varepsilon_{t-1}) = \operatorname{cov}(u_t - (1-\lambda)u_{t-1}; u_{t-1} - (1-\lambda)u_{t-2}) = \operatorname{cov}(-(1-\lambda)u_{t-1}; u_{t-1}) = -(1-\lambda)\operatorname{cov}(u_{t-1}; u_{t-1}) = -(1-\lambda)\sigma_u^2$
Therefore OLS estimators of coefficients will be inconsistent.

Question 7.
A researcher has time series data for aggregate consumption, *C*, and aggregate disposable personal income, *Y*, for a certain country. She establishes that the logarithms of both series are I(1) (integrated of order one). The researcher also believes that $\log C_t$ and $\log Y_t$ are cointegrated, so the cointegrating relationship exists

$$\log C_t = \alpha_1 + \alpha_2 \log Y_t + u_t \qquad (1)$$

where it may be assumed that u_t is normally distributed with zero mean and constant variance.

(a) Explain what is meant when a series is described as being integrated of order one. Explain what is meant when two time series are described as being cointegrated in the context of the problem under consideration. How to test both statements (the series is integrated of order 1, and two series under consideration are cointegrated)? Show that while $\log C_t$ and $\log Y_t$ are not stationary the logarithmic growth rates of consumption $\log \frac{C_t}{C_{t-1}}$ and income $\log \frac{Y_t}{Y_{t-1}}$ are stationary. Can be a regression of one of them on another considered as cointegrating relationship?

Solution
a) The time series is I(1) (integrated of order one) if this time series itself is non-stationary but its difference is stationary. To test this augmented Dickey-Fuller test is applied first to levels of original time series and then to the difference. In both cases the null hypothesis is non-stationarity.
Two time series are cointegrated if there exists a linear combination of the two series that is stable in the sense that the divergences, represented by the disturbance term, are stationary. The researcher should test both time series, their residuals and the difference $\log C_t - \hat{\alpha}_1 - \hat{\alpha}_2 \log Y_t$ for stationarity using a standard unit root test where $\hat{\alpha}_1, \hat{\alpha}_2$ are OLS estimators of the coefficients of regression (1).
The regression (1)

$$\log C_t = \alpha_1 + \alpha_2 \log Y_t + u_t$$

after lagging

$$\log C_{t-1} = \alpha_1 + \alpha_2 \log Y_{t-1} + u_{t-1}$$

and subtracting

$$\log C_t - \log C_{t-1} = \alpha_2(\log Y_t - \log Y_{t-1}) + u_t - u_{t-1}$$

gives

$$\log \frac{C_t}{C_{t-1}} = \alpha_2 \log \frac{Y_t}{Y_{t-1}} + u_t - u_{t-1} \quad (1^*)$$

Both logarithmic growth rates $\log \frac{C_t}{C_{t=1}} = \log C_t - \log C_{t-1} = \Delta \log C_t$ and $\log \frac{Y_t}{Y_{t=1}} = \log Y_t - \log Y_{t-1} = \Delta \log Y_t$ are stationary by assumptions of the model.
The obtained relationship (1*) between logarithmic growth rates cannot be considered as cointegrating relationship as cointegration is not applicable to the stationary time series.

(b) The researcher is also interested in the short-run dynamics of the relationship and correctly hypothesizes that they may be represented by ADL(1,2) process with two lags for income variable

$$\log C_t = \beta_1 + \beta_2 \log C_{t-1} + \beta_3 \log Y_t + \beta_4 \log Y_{t-1} + \beta_5 \log Y_{t-2} + \varepsilon_t \qquad (2)$$

where ε_t is identically and independently distributed and drawn from a normal distribution with zero mean. Derive long run relationship between the equilibrium values of $\log C_t$ and $\log Y_t$: $\overline{\log C_t}$ and $\overline{\log Y_t}$, and find short run and long run income elasticities of consumption.

Solution

Putting $\log C_t = \log C_{t-1} = \overline{\log C}$ and $\log Y_t = \log Y_{t-1} = \log Y_{t-2} = \overline{\log Y}$, (2) implies the long-run relationship

$$\overline{\log C} = \beta_1 + \beta_2 \overline{\log C} + \beta_3 \overline{\log Y} + \beta_4 \overline{\log Y} = \frac{\beta_1}{1-\beta_2} + \frac{\beta_3 + \beta_4 + \beta_5}{1-\beta_2} \overline{\log Y}$$

So β_3 represents short run elasticity while long run elasticity is given by expression $\frac{\beta_3 + \beta_4 + \beta_5}{1-\beta_2}$. Here short run elasticity can be greater or less than long run elasticity depending on the values of the coefficients $\beta_2, \beta_3, \beta_4, \beta_5$.

(c) Using long run relationship between $\log C_{t-1}$ and $\log Y_{t-1}$ as a possible cointegrating relationship derive error-correction type model for the model (2).

Solution

Let us make a plan of transformations needed. From the previous result

$$\overline{\log C} = \frac{\beta_1}{1-\beta_2} + \frac{\beta_3 + \beta_4 + \beta_5}{1-\beta_2} \overline{\log Y}$$

So cointegrating relationship should have the form

$$\log C_{t-1} - \frac{\beta_1}{1-\beta_2} - \frac{\beta_3 + \beta_4 + \beta_5}{1-\beta_2} \log Y_{t-1}$$

From here the error correction style model should be like this

$$\Delta \log C_t = (\beta_2 - 1)\left(\log C_{t-1} - \frac{\beta_1}{1-\beta_2} - \frac{\beta_3 + \beta_4 + \beta_5}{1-\beta_2} \log Y_{t-1} \right) + some \;\; differences + \varepsilon_t$$

Expanding this equation we find

$$\log C_t = \beta_1 + \beta_2 \log C_{t-1} + \underline{\beta_3 \log Y_{t-1}} + \beta_4 \log Y_{t-1} + \underline{\beta_5 \log Y_{t-1}} + some \;\; differences + \varepsilon_t$$

Compare this expression with ADL(1,1) model

$$\log C_t = \beta_1 + \beta_2 \log C_{t-1} + \underline{\underline{\beta_3 \log Y_t}} + \beta_4 \log Y_{t-1} + \underline{\underline{\beta_5 \log Y_{t-2}}} + \varepsilon_t$$

Underlined are expressions that should not be in ADL model – they should be duplicated with the sign 'minus' in difference terms, double underlined expressions, which are lacking for the model – they should be duplicated with the sign 'plus' in difference terms,. Now we can find the differences in the right hand side of the model

$$\Delta \log C_t = (\beta_2 - 1)\left(\log C_{t-1} - \frac{\beta_1}{1-\beta_2} - \frac{\beta_3 + \beta_4 + \beta_5}{1-\beta_2} \log Y_{t-1} \right) +$$
$$+ \beta_3 \log Y_t - \beta_3 \log Y_{t-1} - \beta_5 \log Y_{t-1} + \beta_5 \log Y_{t-2} + \varepsilon_t$$

or finally

$$\Delta \log C_t = (\beta_2 - 1)\left(\log C_{t-1} - \frac{\beta_1}{1-\beta_2} - \frac{\beta_3 + \beta_4 + \beta_5}{1-\beta_2} \log Y_{t-1} \right) + \beta_3 \Delta \log Y_t - \beta_5 \Delta \log Y_{t-1} + \varepsilon_t$$

(d) Describe the structure and dynamics of error correction model. Explain why fitting the error-correction model, rather than (2) directly, avoids a potentially important problem. Which assumptions are necessary to use error correction model? How these assumptions could be tested?

Solution

Any error correction model includes two parts: basic model using first and possibly higher order differences

$$\Delta \log C_t = \beta_3 \Delta \log Y_t - \beta_5 \Delta \log Y_{t-1} + \varepsilon_t$$

and the error correction mechanism

$$(\beta_2 - 1)\left(\log C_{t-1} - \frac{\beta_1}{1-\beta_2} - \frac{\beta_3 + \beta_4 + \beta_5}{1-\beta_2} \log Y_{t-1} \right)$$

based on cointegration relationship. Error correction mechanism allows to compensate partially errors caused by the myopia of the short term relationship described by difference model. The speed of correction depends on the value of the coefficient $\beta_2 - 1$ in the error correction mechanism.

The advantage of fitting the error-correction model, rather than original model

$$\log C_t = \alpha_1 + \alpha_2 \log Y_t + u_t \quad (1)$$

directly, is that the variables in this relationship are all stationary and so the relationship may be fitted with standard OLS. Estimation and further analysis of original model based on non-stationary time series could lead to spurious regression.

The set of tests should be performed: the original time series should be non stationary, their first differences - stationary. The residuals of the regression $\log C_t = \alpha_1 + \alpha_2 \log Y_t + u_t$ should be also stationary).

The error correction model in this case uses first differences, and cointegrating relationship, so all parts of it are stationary.

The International College of Economics and Finance
Econometrics – 2017-2018.
Final exam 2018 May 24.
Suggested solutions

General instructions. Candidates should answer 6 of the following 7 questions: all questions of the Section A and any two of the questions from Section B (questions 5-7). The weight of the Section A is 60% of the exam; two other questions from the Section B add 20% each. You are advised to divide your time accordingly. Structure your answers in accordance with the structure of the questions. When testing hypotheses always state clearly null and alternative hypotheses provide critical value used for test, mentioning degrees of freedom and the significance level chosen for the test.

SECTION A

Answer **ALL** questions from this section (questions **1-4**).

Question 1 The relationship between the US citizens' expenditure on local transport *LOCT* in 1989-2013 (in billions of dollars) and aggregate personal disposable income *DPI* is studied, this relationship is described with the following equation:

$$LOCT_t = 5.31 - 0.28 \ln DPI_t$$
$$(s.e.) \qquad (1.02)\ (0.15)$$

(a) Perform t-tests for significance of both coefficients of the model and give interpretation to them.

For the slope t-statistic is $t = \frac{-0.28}{0.15} = -1.87$ while $t_{crit}(5\%, 23) = 2.069$, so the slope coefficient is insignificant; for intercept $t = \frac{5.31}{1.02} = 5.21$ while $t_{crit}(1\%, 23) = 2.807$ so it is significant at 1% level.
Formally the slope coefficient -0.28 shows that if disposable personal income increases by 1% the expenditure on local transport decreases by $\frac{0.28}{100} = 0.0028\%$. But it is insignificant, so for the equation

$$LOCT_t = \beta_1 + \beta_2 \ln DPI_t + u_t$$

the null hypothesis $H_0 : \beta_2 = 0$ is not rejected, so the equation reduces to $LOCT_t = \beta_1 + u_t$ where β_1 can be interpreted as the average expenditures on local transport independently of DPI_t.

(b) In the equation above, there is no information about the R-square. Is it possible to restore its value from the available information? Is it high enough to make the equation significant?

For the simple linear regression t-test and F-test are equivalent so according F-test the equation is olso insignificant. Moreover $F = t^2 = (1.87)^2 = 3.5$, so solving equation $\frac{R^2}{1-R^2}(25-2) = 3.5$ find $R^2 = 0.132$.

(c) At the seminar one of the participants remarked that coefficient of $\log DPI_t$ is expected to be negative as people tend to use personal cars instead of local transport as their income rises. The other participant objected to him saying that just the opposite is true (the coefficient of $\log DPI_t$ should be positive) as local transport is used mainly by elderly people who prefer not to use cars and are quite sensitive to changes in income. How both suggestions change your conclusion on significance of the coefficients of the model?
Answering all questions state clearly null and alternative hypotheses, degrees of freedom, used critical values.

If the pair of hypotheses $H_0 : \beta_2 = 0; H_a : \beta_2 < 0$ is used with one sided critical value $t_{crit}(5\%, 23, one\ sided) = -1.714$ the slope coefficient becomes significant. If the pair of hypotheses $H_0 : \beta_2 = 0; H_a : \beta_2 > 0$ is used the coefficient is obviously insignificant because the observed value contradicts the alternative.

Question 3 The simple linear regression $Y_i = \beta_1 + \beta_2 X_i + u_i$, $i = 1, 2, ..., n$ is considered. $\hat{Y}_i = b_1 + b_2 X_i$ are estimated values of dependent variable, the residuals are $\hat{u}_i = Y_i - b_1 - b_2 X_i$ where b_1 and b_2 are ordinary least squares estimates of β_1 and β_2.

(a) Show that $\mathrm{Var}(Y_i) = \mathrm{Var}(\hat{Y}_i) + \mathrm{Var}(\hat{u}_i)$, explaining clearly how this equality follows from OLS principle.

$Y_i = \hat{Y}_i + \hat{u}_i$ hence $\mathrm{Var}(Y_i) = \mathrm{Var}(\hat{Y}_i) + \mathrm{Var}(\hat{u}_i) + 2\,\mathrm{Cov}(\hat{Y}_i, \hat{u}_i)$. Now,
$\mathrm{Cov}(\hat{Y}_i, \hat{u}_i) = \mathrm{Cov}(b_1 + b_2 X_i, \hat{u}_i) = \mathrm{Cov}(b_1, \hat{u}_i) + b_2\,\mathrm{Cov}(X_i, \hat{u}_i)$, which is 0 since $\mathrm{Cov}(b_1, \hat{u}_i) = 0$ as b_1 is a constant and $\mathrm{Cov}(X_i, \hat{u}_i) = 0$ by the normal equations of OLS (see below *).

Hence $\mathrm{Var}(Y_i) = \mathrm{Var}(\hat{Y}_i) + \mathrm{Var}(\hat{u}_i)$.

[Note: Variances and covariances are sample variances and sample covariances.]

*) By definition $\mathrm{Cov}(X_i, \hat{u}_i) = \frac{1}{n}\left(\sum X_i \cdot \hat{u}_i\right) - \frac{1}{n}\left(\sum X_i\right) \cdot \frac{1}{n}\left(\sum \hat{u}_i\right) = \frac{1}{n}\left(\sum X_i \hat{u}_i\right)$ as $\sum \hat{u}_i = 0$, so
$\mathrm{Cov}(X_i, \hat{u}_i) = \frac{1}{n}(\sum X_i \hat{u}_i) = \frac{1}{n}(\sum X_i \cdot (Y_i - b_1 - b_2 X_i)) = 0$ as $\sum X_i Y_i = b_1 \sum X_i + b_2 \sum X_i Y_i$ is the second normal equation of OLS.

Alternative solution.

$$TSS = \sum (Y_i - \bar{Y})^2 = \sum ((\hat{Y}_i + \hat{u}_i) - \bar{Y})^2 = \sum ((\hat{Y}_i - \bar{Y}) + \hat{u}_i)^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum \hat{u}_i^2 + 2\sum (\hat{Y}_i - \bar{Y})\hat{u}_i =$$
$$= \sum (\hat{Y}_i - \bar{Y})^2 + \sum \hat{u}_i^2 + 2\sum \hat{Y}_i \hat{u}_i + 2\bar{Y}\sum \hat{u}_i = \sum (\hat{Y}_i - \bar{Y})^2 + \sum \hat{u}_i^2 = ESS + RSS$$

as $\sum \hat{u}_i = 0$ and $\sum \hat{Y}_i \hat{u}_i = \sum (\hat{\beta}_1 + \hat{\beta}_2 X_i)\hat{u}_i = \hat{\beta}_1 \sum \hat{u}_i + \hat{\beta}_2 \sum X_i \hat{u}_i = 0 + 0 = 0$

So $\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum \hat{u}_i^2$

Dividing this expression by n and taking into account that $\bar{\hat{Y}} = \bar{Y}$ and $\bar{u} = 0$ we get

$$\frac{1}{n}\sum (Y_i - \bar{Y})^2 = \frac{1}{n}\sum (\hat{Y}_i - \bar{Y})^2 + \frac{1}{n}\sum \hat{u}_i^2 \quad \text{or} \quad \mathrm{Var}(Y_t) = \mathrm{Var}(\hat{Y}_t) + \mathrm{Var}(\hat{u}_t)$$

(by definition $\mathrm{Var}(X) = \frac{1}{n}\sum (X_i - \bar{X})^2$)

(b) Explain how this expression is related to the properties of R^2, the coefficient of determination. Comment on its meaning and its role in regression analysis.

Multiplying by T gives $TSS = ESS + RSS$. Now $R^2 = 1 - RSS/TSS = ESS/TSS$ as $R^2 \geq 0$ since ESS and TSS must be positive and $TSS > 0$. Also $R^2 \leq 1$ since $ESS \leq TSS$ as $RSS \geq 0$.

According to the definition $R^2 = \frac{ESS}{TSS}$ the determination coefficient shows the proportion of variance of dependent variable 'explained' by the regression equation. The closer it to unity, in general, the better the regression equation. This property is especially important for forecasting.

R^2 is equal to the square of the correlation coefficient between X_i and Y_i.

There are no special tables to determine R^2 to be sufficiently close to one. To test the quality of the regression equations, an F-test is usually used: $F = \frac{R^2}{1 - R^2} \cdot (n-2)$ having F-distribution with $(1, n-2)$ degrees of freedom.

(c) **i)** Show that $\mathrm{Cov}(Y_i, \hat{Y}_i) = \mathrm{Var}(\hat{Y}_i)$

By definition $\text{Cov}(Y_i,\hat{Y}_i)=\frac{1}{n}\sum(Y_i-\bar{Y})(\hat{Y}_i-\bar{Y})$ and $\text{Var}(\hat{Y}_i)=\frac{1}{n}\sum(\hat{Y}_i-\bar{\hat{Y}})^2=\frac{1}{n}\sum(\hat{Y}_i-\bar{Y})^2$ as $\bar{\hat{Y}}=\bar{Y}$. So we have to show that $\sum(Y_i-\bar{Y})(\hat{Y}_i-\bar{Y})=\sum(\hat{Y}_i-\bar{Y})^2$.

$$\sum(Y_i-\bar{Y})(\hat{Y}_i-\bar{Y})=\sum((\hat{Y}_i+\hat{u}_i)-\bar{Y})(\hat{Y}_i-\bar{Y})=\sum((\hat{Y}_i-\bar{Y})+\hat{u}_i)(\hat{Y}_i-\bar{Y})=$$
$$=\sum(\hat{Y}_i-\bar{Y})^2+\sum\hat{u}_i\hat{Y}_i-\bar{Y}\sum\hat{u}_i=\sum(\hat{Y}_i-\bar{Y})^2$$

as $\sum\hat{u}_i=0$ and $\sum\hat{Y}_i\hat{u}_i=\sum(\hat{\beta}_1+\hat{\beta}_2X_i)\hat{u}_i=\hat{\beta}_1\sum\hat{u}_i+\hat{\beta}_2\sum X_i\hat{u}_i=0+0=0$

Alternative solution.

$$\text{Cov}(Y_i,\hat{Y}_i)=\text{Cov}((\hat{Y}_i+\hat{u}_i),\hat{Y}_i)=\text{Cov}(\hat{Y}_i,\hat{Y}_i)+\text{Cov}(\hat{u}_i,\hat{Y}_i)=\text{Var}(\hat{Y}_i)+0=\text{Var}(\hat{Y}_i)$$

as $\text{Cov}(\hat{u}_i,\hat{Y}_i)=\frac{1}{n}\sum\hat{Y}_i\hat{u}_i=\frac{1}{n}\sum(b_1+b_2X_i)\hat{u}_i=\frac{1}{n}(b_1\sum\hat{u}_i+b_2\sum X_i\hat{u}_i)=0+0=0$

ii) Show that the correlation between actual and fitted values of dependent variable is always positive and is equal to the square root of the determination coefficient: $r_{Y,\hat{Y}}=\sqrt{R^2}$.

$$r_{Y,\hat{Y}}=\frac{\sum(Y_i-\bar{Y})(\hat{Y}_i-\bar{Y})}{\sqrt{\sum(Y_i-\bar{Y})^2\sum(\hat{Y}_i-\bar{Y})^2}}=\frac{\sum(\hat{Y}_i-\bar{Y})^2}{\sqrt{\sum(Y_i-\bar{Y})^2\sum(\hat{Y}_i-\bar{Y})^2}}=\frac{\sqrt{\sum(\hat{Y}_i-\bar{Y})^2}}{\sqrt{\sum(Y_i-\bar{Y})^2}}=\sqrt{\frac{\sum(\hat{Y}_i-\bar{Y})^2}{\sum(Y_i-\bar{Y})^2}}=\sqrt{R^2}$$

Alternative solution.

$$r_{Y,\hat{Y}}=\frac{\text{Cov}(Y_i,\hat{Y}_i)}{\sqrt{\text{Var}(Y_i)\cdot\text{Var}(\hat{Y}_i)}}=\frac{\text{Var}(\hat{Y}_i)}{\sqrt{\text{Var}(Y_i)\cdot\text{Var}(\hat{Y}_i)}}=\frac{\sqrt{\text{Var}(\hat{Y}_i)}}{\sqrt{\text{Var}(Y_i)}}=\sqrt{\frac{\text{Var}(\hat{Y}_i)}{\text{Var}(Y_i)}}=\sqrt{R^2}$$

Question 4 A researcher wants to examine the newspaper reading habits of households. For this she collects data on fifty households and defines

$y_i=1$ if the i-th household purchases a newspaper

$\quad =0$ otherwise.

She estimates the linear regression model defining $y_i = f(S_i, E_i, u_i)$ where

S_i = years spent by the head of the i-th household in full time education

E_i = average earnings of the head of the i-th household

u_i = unobserved disturbance term.

The model was estimated by ordinary least squares (LPM – linear probability model) and logit with the following results:

	LPM	Logit
S	0.099	0.521
	(4.07)	(3.10)
E	0.012	0.067
	(2.29)	(1.84)
Constant	0.015	–2.56
	(0.16)	(–3.57)

the figures in parentheses are the t values (in LPM model) and asymptotic z-values in logit model.

(a) What is the difference in estimation of LPM and logit models? What is the difference in imerpretation of coefficients of LPM and logit models? What are the comparative advantages of the logit model in front of a linear probability model, estimated by the least squares?

Linear probability model is estimated by OLS, while logit model is estimated using Maximum Likelihood Principle. In both models estimated values of dependent variable are interpreted as the probability of purchasing a newspaper, but only logit model guarantees that this probability will lie in the range from 0 to 1. Another advantage of the logit model is the variable marginal effects of both explanatory variables, whereas in the LPM they are constant. Both models suffer from the presence of heteroscedasticity, but in the logit model this is not so important, since the maximum likelihood method used to estimate the logit model is not very sensitive to violation of Gauss-Markov conditions if the sample is large enough. Another feature of both models is the distribution of the random term is not normal due to the fact that the dependent variable takes only two values (0 and 1). This invalidates all tests in LPM, whereas logit model uses quite different set of tests based on the likelihood function.

(b) Obtain the predicted probability for the i-th household if $S_i = 7$ and $E_i = 40$ from both sets of estimates.

LPM

$$\hat{Y}_i = 0.015 + 0.099(7) + 0.012(40) = 1.188$$

$$P(Y_i = 1) = 1.188$$

So LPM model gives estimate of probability outside the range [0; 1].

Logit

$$\hat{P}_i = P(Y_i = 1) = \frac{\exp(\hat{Y}_i)}{1 + \exp(\hat{Y}_i)}$$

$$\hat{Y}_i = -2.56 + 0.521(7) + 0.067(40) = 3.767$$

$$\hat{P}_i = P(Y_i = 1) = \frac{\exp(3.767)}{1 + \exp(3.767)} = 0.977$$

As it follows from the theory this result is in the admissible range [0; 1].

(c) Evaluate the marginal effect of average earnings E_i for both models in the same point $S_i = 7$ and $E_i = 40$. Are these effects significant? Compare evaluated marginal effect for logit model with the maximum one.

In LPM the coefficients show the constant marginal effects of factors, so in the point All the slope coefficients are significant. $\hat{P}_i = P(Y_i = 1) = \hat{Y}_i$ $S_i = 7$ and $E_i = 40$ the effects are 0.099 for S_i and 0.012 for E_i.

To evaluate the marginal effect of the factor S_i at the point $S_i = 7$ and $E_i = 40$, we have to multiply the value of the derivative of the logit function at this point by the coefficient of S_i in logit regression $\frac{\exp(-3.767)}{(1+\exp(-3.767))^2} \cdot 0.521 = 0.0115$. This value lies far away from the maximum effect that is achieved at point 0 - the center of symmetry of the derivative of the logistic curve: $\frac{\exp(0)}{(1+\exp(0))^2} \cdot 0.521 = 0.13$.

Alternative solution.

It is possible to evaluate marginal effect without derivative. Let S_i changes from 7 to 8 keeping $E_i = 40$. Then

$\hat{Y}_i(7, 40) = -2.56 + 0.521(7) + 0.067(40) = 3.767$ and $\hat{Y}_i(8, 40) = -2.56 + 0.521(8) + 0.067(40) = 4.288$

$\Delta \hat{P}_i = \frac{\exp(4.288)}{1+\exp(4.288)} - \frac{\exp(3.767)}{1+\exp(3.767)} = 0.009$ what is close to the estimate 0.0115 of derivative method.

SECTION B

Answer **TWO** questions from 5-7.

Question 5

1. Consider the model:

$$y_t = ax_t + u_t\,,\ t = 1, \ldots, T$$

where $E(u_t) = 0$, $E(u_t^2) = \sigma^2$, $E(u_t u_{t'}) = 0$ for $t \neq t'$.

Suppose that the parameter a changes at a certain point s the sample, i.e.

$$a = a^*\,,\ t = 1, \ldots, s$$
$$a = a^{**}\,,\ t = s+1, \ldots, T$$

(a) Explain how you would estimate a^* and a^{**}:

(i) not using dummy variables;

(ii) using dummy variables.

(i) Run regression $y_t = ax_t + u_t$ using two samples $t = 1, 2, \ldots, s$ and $t = 1, 2, \ldots, s$.

(ii) First introduce dummy variable $d_t = 0$ for $t = 1, 2, \ldots, s$ and $d_t = 1$ for $t = s+1, \ldots, n$.

Specify the model as $y_t = (a + \lambda d_t)x_t + u_t = ax_t + \lambda(d_t x_t) + u_t$ for $t = 1, 2, \ldots, T$ Apply OLS to this model – the estimate $\hat{a}$ is an estimate of a^* and $\hat{a} + \hat{\lambda}$ is an estimate of a^{**}.

(b) Suppose an econometrician ignores the change in a, and assumes that a is constant throughout the sample. Derive an expression for the bias of the OLS estimate of a as an estimate of a^*.

<u>**Solution and marking**</u>

$$\hat{a} = \frac{\sum x_t y_t}{\sum x_t^2} = \frac{\sum x_t(ax_t + \lambda d_t x_t + u_t)}{\sum x_t^2} = a\frac{\sum x_t x_t}{\sum x_t^2} + \lambda\frac{\sum d_t x_t x_t}{\sum x_t^2} + \frac{\sum x_t u_t}{\sum x_t^2} = a + \lambda\frac{\sum_{t=1}^{s} x_t^2}{\sum_{t=1}^{T} x_t^2} + \frac{\sum x_t u_t}{\sum x_t^2}$$

(here the sum $\sum$ means $\sum_{t=1}^{T}$, unless otherwise indicated).

so $$E(\hat{a}) = a + \lambda\frac{\sum_{t=1}^{s} x_t^2}{\sum_{t=1}^{T} x_t^2} + \frac{\sum x_t E(u_t)}{\sum x_t^2} = a + \lambda\frac{\sum_{t=1}^{s} x_t^2}{\sum_{t=1}^{T} x_t^2}$$ since $E(u_t) = 0$. The bias is therefore $\lambda\frac{\sum_{t=1}^{s} x_t^2}{\sum_{t=1}^{n} x_t^2} = \frac{\lambda\sum d_t x_t^2}{\sum x_t^2}$.

(c) Given data on x_t and y_t explain how you would test your model for a change in the slope coefficient at time s

(i) using a t-test;

(ii) using F-test for restriction (what is the restriction?).

(i) Estimate the equation with the dummy variable $y_t = ax_t + \lambda(d_t x_t) + u_t$ for $t = 1, 2, \ldots, T$ and use a *t*-test for the estimate of λ: $t = \dfrac{\hat{\lambda}}{s.e.(\hat{\lambda})}$ and compare it with $t_{crit}(\alpha, T-2)$.

(ii) Estimate the model as $y_t = ax_t + u_t$ for $t = 1, 2, \ldots, T$ and memorize restricted sum of squares RSS_R. Then estimate the model as $y_t = ax_t + \lambda(d_t x_t) + u_t$ for $t = 1, 2, \ldots, T$ and memorize unrestricted sum of squares RSS_U. Now evaluate F-test for restriction $\lambda = 0$: $F_{1,T-2} = \dfrac{(RSS_R - RSS_U)/1}{RSS_U/(T-2)}$. Of course $F = t^2$ for t-statistic obtained in (i).

(d) An alternative to the test in (c) is a Chow test. Explain how you would apply a Chow test to this model and how the results would compare with the test you described in (c).

Estimate the equation $y_t = ax_t + ut$ for $t = 1, 2, ..., s$ to give RSS_1. Estimate the same equation for $t = s+1, ..., n$ to give RSS_2. Now estimate the same equation on the whole sample, i.e. $t = 1, 2, ..., T$ to give RSS_R. Now $RSS_U = RSS_1 + RSS_2$ and the Chow test becomes $F_{1,T-2} = \dfrac{(RSS_R - RSS_U)/1}{RSS_U/(T-2)}$. Of course it gives the same value as F-test in **(c).**

Question 6. Consider a model

$$y_t = \theta y_{t-1} + u_t;\ t = 1, 2, ..., T$$

where $y_0 = 0$. $E(u_t) = 0$, $E(u_t^2) = \sigma^2$ and $E(u_s u_t) = 0$ when $s \neq t$, for all $s, t = 1, 2, ..., T$.

(a) Derive the mean and variance of y_t when $\theta = 1$ and comment on the result.

The model is (random walk):

$$Y_t = Y_{t-1} + u_t,\ t = 1, 2, ..., T.$$

We can write:

$$t = 1,\ Y_1 = u_1,$$
$$t = 2,\ Y_2 = Y_1 + u_2 = u_1 + u_2,$$
$$t = 3,\ Y_3 = Y_2 + u_3 = u_1 + u_2 + u_3$$
$$\vdots, \qquad \vdots$$

Doing these recursive substitutions, we can write:

$$Y_t = Y_{t-1} + u_t = Y_0 + u_t + u_{t-1} + ... + u_1 = Y_0 + \sum_{s=1}^{t} u_s,$$

where for simplicity Y_0 may be supposed fixed.

So:

$$E(Y_t) = Y_0 + \sum_{s=1}^{t} Eu_s = Y_0 \text{ (constant in } t\text{)}$$

but:

$$\mathrm{Var}(Y_t) = \mathrm{Var}\left(\sum_{s=1}^{t} u_s\right) = t\sigma^2.$$

Thus Y_t is non-stationary as the variance of Y_t is dependent on time.

(b) Describe in detail the Dickey-Fuller and the augumented Dickey-Fuller (ADF) procedure for testing for the order of integration of a time series variable.

The standard test for a unit root is due to Dickey and Fuller and is based on the model:

$$y_t = \beta_1 + \beta_2 y_{t-1} + \gamma t + u_t,$$

which can be re-written as:

$$\Delta y_t = \beta_1 + (1 - \beta_2) y_{t-1} + \gamma t + u_t,$$

where $\Delta y_t = y_t - y_{t-1}$.

The null hypothesis for stationarity is:

$$H_0: 1 - \beta_2 = 0,\ H_A: 1 - \beta_2 \neq 0.$$

We cannot use the standard *t*-test procedure in this case because the distribution of the *t*-statistic is not a *t*-distribution, so critical values have been computed by Dickey and Fuller using Monte-Carlo techniques.

The test is sensitive to the presence of serial correlation in the error term so we need to take steps to remove the effects of this serial correlation - this is done by including lagged values of y_t in the regression, i.e.:

$$y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \gamma_t t + u_t$$

for an AR(1) serial correlation. To derive Dickey-Fuller equation we have first to subtract lagged value of y_t from both sides of equation

$$y_t - y_{t-1} = \beta_1 + \beta_2 y_{t-1} - y_{t-1} + \beta_3 y_{t-2} + \gamma_t t + u_t$$

Then we have to add and simultaneously subtract in right hand side the term $\beta_3 y_{t-1}$ as both sides of equation should contain only stationary differences

$$y_t - y_{t-1} = \beta_1 + \beta_2 y_{t-1} - y_{t-1} + (\beta_3 y_{t-1} - \beta_3 y_{t-1}) + \beta_3 y_{t-2} + \gamma_t t + u_t$$

$$y_t - y_{t-1} = \beta_1 + \beta_2 y_{t-1} + \beta_3 y_{t-1} - y_{t-1} - (\beta_3 y_{t-1} - \beta_3 y_{t-2}) + \gamma_t t + u_t$$

or finally

$$\Delta y_t = \beta_1 + (\beta_2 + \beta_3 - 1) y_{t-1} - \beta_3 \Delta y_{t-1} + \gamma_t t + u_t ,$$

with null hypothesis $H_0 : 1 - \beta_1 - \beta_2 = 0$. To test for unit root we test the coefficient of Y_{t-1}, i.e.:

$$H_0 : \beta_2 + \beta_3 - 1 = 0 ,\ H_1 : \beta_2 + \beta_3 - 1 < 0 .$$

Once again, the Dickey-Fuller tables should be used.

(c) Consider a model

$$Y_t = \alpha_1 + \alpha_2 Y_{t-1} + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

where $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma^2$ and $E(\varepsilon_s \varepsilon_t) = 0$ for all $s \neq t,\ \ s, t = 1, 2, ..., T$. Derive the specification for the ADF test.

The model is:

$$Y_t = \alpha_1 + \alpha_2 Y_{t-1} + u_t ,\ u_t = \rho u_{t-1} + \varepsilon_t ,$$

where ε_2 is $I(0)$. Dickey-Fuller test is sensitive to autocorrelation so we should remove it first. We know that autocorrelation of this type can be eliminated or significantly reduced efficiently by autoregression transformation. So lag this equation by one period and multiply it by ρ to get:

$$\rho Y_{t-1} = \alpha_1 \rho + \alpha_2 \rho Y_{t-2} + \rho u_{t-1} .$$

Subtract it from original equation and rearrange to get:

$$Y_t = \alpha_1 (1 - \rho) + (\alpha_2 + \rho) Y_{t-1} - \alpha_2 \rho Y_{t-2} + (u_t - \rho u_{t-1}) =$$

$$= \alpha_1 (1 - \rho) + (\alpha_2 + \rho) Y_{t-1} - \alpha_2 \rho Y_{t-2} + \varepsilon_t .$$

This can be written as:

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 Y_{t-2} + \varepsilon_t ,$$

where $\beta_1 = \alpha_1 (1 - \rho)$, $\beta_2 = (\alpha_2 + \rho)$, $\beta_3 = -\alpha_2 \rho Y_{t-2}$.

The model of this type was discussed in (b), so we get Dickey-Fuller equation

$$\Delta y_t = \beta_1 + (1 - \beta_2 - \beta_3) y_{t-1} - \beta_3 \Delta y_{t-1} + \varepsilon_t ,$$

To test for unit root we test the coefficient of Y_{t-1}, i.e.:

$$H_0 : \beta_2 + \beta_3 - 1 = 0 ,\ H_1 : \beta_2 + \beta_3 - 1 < 0 .$$

Once again, the Dickey-Fuller tables should be used.

(d) Consider a model

$$Y_t = \beta X_t + u_t ;\ t = 1, 2, ..., T$$

$$u_t = \theta e_{t-1} + e_t$$

where $E(e_t) = 0$, $E(e_t^2) = \sigma^2$ and $E(e_s e_t) = 0$ if $s \neq t$ for all $s, t = 1, 2, ..., T$.

Are Y_t and X_t cointegrated? Explain.

Y_t and X_t are cointegrated if a linear combination of Y_t and X_t is $I(0)$. $u_t = Y_t - \beta X_t$ is a linear combination of Y_t and X_t. Hence, we have to examine the stationarity of u_t:

$$E(u_t) = E(e_t + \theta e_{t-1}) = Ee_t + \theta Ee_{t-1} = 0.$$

$$E(u_t^2) = E(e_t^2) + \theta^2 E(e_{t-1}^2) + 2E(e_t e_{t-1}) = (1+\theta^2)\sigma^2,$$

since $E(e_t e_{t-1}) = 0$.

$$E(u_t u_{t-1}) = E(e_t + \theta e_{t-1})(e_{t-1} + \theta e_{t-2}) =$$
$$= E(e_t e_{t-1}) + \theta E(e_t e_{t-2}) + \theta E(e_{t-1}^2) + \theta^2 E(e_{t-1} e_{t-2}) = \theta\sigma^2,$$

since $E(e_t e_{t-s}) = 0$, for all $s > 0$.

$$E(u_t u_{t-2}) = E(e_t + \theta e_{t-1})(e_{t-2} + \theta e_{t-3}) =$$
$$= E(e_t e_{t-2}) + \theta E(e_t e_{t-3}) + \theta E(e_{t-1} e_{t-2}) + \theta^2 E(e_{t-1} e_{t-3}) = 0.$$

Thus both first and second moments are independent of *t* and u_t must be (weakly) stationary. This implies that Y_t and X_t are cointegrated.

Question 7 A student evaluates logarithmic regression of the total expenditures of USA citizens on tobacco *TOB* (billions of dollars) on price index *PTOB* for the period 1959-1983 (100% corresponds to the level of the year 1972).

$$\underset{(0.09)\ (0.02)}{\ln \hat{TOB}_t = 1.56 + 0.21 \ln PTOB_t} \qquad \begin{matrix} R^2 = 0.82 \\ d = 0.76 \end{matrix} \quad (1)$$

Trying to remove autocorrelation the student applies autoregressive transformation AR(1)

$$\underset{(4.84)\ (0.20)}{\ln \hat{TOB}_t = 5.18 - 0.28 \ln PTOB_t} \qquad \begin{matrix} R^2 = 0.92 \\ d = 2.61 \end{matrix} \quad (2)$$

(a) Give interpretation of the coefficients of this equation (1). Find some indications of the autocorrelation in this equation. What is the autoregressive transformation AR(1), show mathematically how autoregressive transformation works? Was this transformation successful in removing autocorrelation (equation 2)?

The coefficient 0.21 can be interpreted as price elasticity of tobacco consumption, it means that With an increase in the price index by one percentage point the consumption of tobacco rises by 0.21%. Although this coefficient is significant, the absence of important explanatory variables and economic considerations make it possible to say that this equation is unsatisfactory. Durbin-Watson statistic equal to 0.76 indicates the presence of autocorrelation: $0.76 < 1.05 = d_{crit}(1\%, 25, \textit{number of parameters} = 2, \textit{lower})$.

Assuming 1-st order autocorrelation $u_t = \rho u_{t-1} + \varepsilon_t$, where $E\varepsilon_t = 0$, $E(\varepsilon_t^2) = \sigma_\varepsilon^2$, $E(\varepsilon_t \varepsilon_s) = 0, s \neq t$, we can subtract from initial equation

$$\ln TOB_t = \beta_1 + \beta_2 PTOB_t + u_t$$

the lagged equation multiplied by ρ:

$$\rho \ln TOB_{t-1} = \rho\beta_1 + \beta_2 \rho PTOB_{t-1} + \rho u_{t-1}$$

If ρ is taken correctly the disturbance term becomes $u_t - \rho u_{t-1} = \varepsilon_t$:

$$\ln TOB_t - \rho \ln TOB_{t-1} = \beta_1(1-\rho) + \beta_2(PTOB_t - \rho PTOB_{t-1}) + \varepsilon_t.$$

otherwise the autoregressive transformation can be applied repeatedly, refining each time the value of ρ from the Darbin-Watson statistics of the preceding equation $\hat{\rho} = 1 - d/2$.

After repeating autoregressive transformation $d = 2.61$, what indicates the negative autocorrelation. So making additional transformation $4 - d = 4 - 2.61 = 1.39$ use comparison with the critical value $d_{crit}(5\%, 25, \textit{number of param.} = 2, \textit{lower}) = 1.30 < 1.39 < 1.46 = d_{crit}(5\%, 25, \textit{number of param.} = 2, \textit{upper})$ – dark zone, no conclusion.

(b) The teacher pointed out that autoregressive transformation made the slope coefficient insignificant and advised the student to include instead into equation more variables: $\ln DPI_t$ − logarithm of the gross disposable personal income (billions of dollars) and $TIME_t$ (which is equal 1 in 1959):

$$\begin{array}{ll}\ln \hat{TOB}_t = \underset{(1.74)}{7.23} - \underset{(0.11)}{0.37}\ln PTOB_t + \underset{(0.22)}{0.55}\ln DPI_t + \underset{(0.012)}{0.049}TIME_t & R^2 = 0.93 \\ & d = 1.26\end{array} \quad (3)$$

Give interpretation of the coefficients of this equation (3). Explain why this removes partially the autocorrelation problem. Was it removed fully or not?

The coefficient -0.37 is price elasticity of tobacco consumption evaluated under assumption of constant income and time, now it has correct sign. The coefficient 0.049 can be interpreted as follows:
be interpreted as price elasticity of tobacco consumption, it means that each year (keeping prices and income constant) the consumption of tobacco increases by $0.049 \cdot 100 = 4.9$ percent.
One of the reasons for the observed autocorrelation may be the incorrect specification of equation. If an important variable is omitted in the equation, the values of which depend on its past values, then this variable, entering into the disturbance term of the equation, can cause the autocorrelation of the disturbance term. Therefore, the inclusion of missing variables can help to cope with the problem of autocorrelation.
$d_{crit}(5\%, 25, \textit{number of param.} = 4\ \textit{lower}) = 1.10 < 1.26 < 1.66 = d_{crit}(5\%, 25, \textit{number of param.} = 4, \textit{upper})$, As the Durbin-Watson statistics in equation (3) is in the dark zone, it is not possible to state autocorrelation.

(c) Additionally the student decided to do Breusch-Godfrey test and in response to the command AUTO(1) she got the value of **Obs*R-squared** equal to 3.3087773 . Explain what is Breusch-Godfrey test and how it works. What is R-squared here? Help the student interpret the test results.

Breusch-Godfrey test involves the evaluation of the auxiliary equation

$$. \ln TOB_t = \beta_1 + \beta_2 PTOB_t + \beta_3 \ln DPI_t + \beta_4 TIME_t + RESID_{t-1} + u_t$$

Where $RESID_{t-1}$ are lagged residuals of equation (3). Under null hypothesis of no autocorrelation the statistics $T \cdot R^2$ (T is the number of observation, R^2 is R-squared of auxiliary equation) has χ^2-distribution with 1 degree of freedom (the number of lagged residuals in the auxiliary equation – here 1 according to the command AUTO(1)). As $\chi^2_{crit}(5\%, df = 1) = 3.8415 > 3.3088$ the null hypothesis of no autocorrelation cannot be rejected.

(d) A student friend said that tobacco consumption can be explained mainly by the habit of smoking. So she tries one more equation with the lagged variable

$$\begin{array}{ll}\ln \hat{TOB}_t = \underset{(0.16)}{0.25} + \underset{(0.06)}{0.90}\ln TOB_{t-1} & R^2 = 0.90 \\ & d = 2.17\end{array} \quad (4)$$

Did a friend's advice help get rid of autocorrelation?

One should use h-statistics ($h = (1-d/2)\sqrt{\dfrac{n}{1-n\cdot \mathrm{Var}(b)}}$) which is equal here $h = (1-2.17)\sqrt{\dfrac{24}{1-24\cdot(0.06)^2}} = -0.435$. Using normal distribution $z_{crit}(5\%) = -1.96$ we come to the conclusion that hypothesis of no autocorrelation cannot be rejected. The slope coefficient is significant so the idea of strong influence of habits on smoking looks reasonable.

The International College of Economics and Finance
Econometrics – 2010-2011.
Mid-year exam. December 27.
Marking scheme.
(This is final version of the marking scheme and not the solution!!!)

Part 2. (1 hour 30 minutes). Answer the first (obligatory) problem and any one of the two problems (2-3). All questions of the first part bear identical number of marks.

1.. (a) An economist believes that the variable Y is determined by two explanatory variables X_2 and X_3 according to the multiple linear regression:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

She estimated the parameters of this equation using a sample of 13 observations (standard errors in parentheses)

$$Y_i = -16.5 + 1.4X_{2i} + 0.8X_{3i} + u_i, \ R^2 = 0.47$$
$$(20.2) \quad (0.7) \quad\quad (0.6)$$

(a1) How the economist should test
1) is β_2 different from zero ,
2) is β_3 different from zero

if she does not know is the influence of both explanatoty variables positive or negative?

(a2) How the economist should test jointly β_2 and β_3 are significantly different from zero.

(a3) How the results of the tests and the conclusions would change if the researcher suggests that X_{2i} and X_{3i} could not influence positively the dependent variable Y_i.

Perform the appropriate tests and discuss the issue of model specification.

Principles of marking and comments

a1)

t-statistics for the slope coefficients are 1.99 and 1.34 correspondingly, while two-sided critical value for 5% is 2.16 so both coefficients are insignificant. **[5 marks]**

a2) $H_0: \beta_2 = 0, \beta_3 = 0$ against alternative 'at least one of the slope coefficients is zero' is rejected at 5% significance level (only) as $F = \dfrac{R^2/2}{(1-R^2)/(13-3)} = \dfrac{0.473/2}{(1-0.473)/10} = 4.488$ what is greater than 4.1. **[5 marks]**

a3) The researcher suggests that X_{2i} and X_{3i} could not influence positively the dependent variable Y_i but the actual values of slope coefficients are positive so one-sided tests would not help here. Possibly the researcher is not right in her suggestions or maybe the sample is too small. In any case the significance of the equation as a whole and insignificance of both coefficients suggest the presence of multicollinearity. **[5 marks]**
[Total 15 marks for a)].

(b) A marketing researcher suspects that Champagne sales increase in a month before Christmas and New Year celebration. She has monthly data for some years from the big retail network having the shops both in big cities, and in the countryside in various regions differing with the level of incomes. How should she use dummy variables to model this data if she supposes that the sales volume S_i depends on the level of income in the region INC, location of the shop L (big city or the countryside) and of time of the year (is this month December or not) D? She also suggests that marginal effect of income is different before Christmas and in other months, but doesn't depend on the location of shop, she also believes that sales change in time differently in big cities and in the countryside. *(For those who are familiar with panel data model: ignore all aspects connected with panel data model estimation, and consider pooled regression)*

(b1) How the dummy variables D and L could be defined? How many other dummy variables should be introduced to satisfy to all assumptions of the researcher, and how it could be done?

(b2) Write down the linear regression model including all dummy variables defined in (b1).

(b3) Give interpretation to the coefficients of regression (*one for each type of variables would be enough*).

Principles of marking and comments

b1) To take into account these factors one should introduce several dummy variables. First of all one define $D = \begin{cases} 1 \text{ if } December \\ 0 \text{ if } other\ months \end{cases}$ and $L = \begin{cases} 1 \text{ if } the\ location\ is\ in\ a\ big\ city \\ 0 \text{ if } the\ location\ is\ in\ a\ country \end{cases}$. Then one define the slope dummy $DINC = D * INC$ and interactive dummy $DL = D * L$. **[5 marks]**

b2) $S_i = \beta_1 + \beta_2 D_i + \beta_3 L_i + \beta_4 INC_i + \beta_5 DINC_i + \beta_6 DL_i + u_i$ **[5 marks]**

b3) In this equation β_1 shows the level of Champagne sales in a country in all months except December by a person of zero registered income (not too informative from economic point of view). β_2 shows the increase of the Champagne sales for this person in December while β_3 shows the increase of the Champagne sales for the person with **any** income living in the big city by comparison of the same person living in a country. Meaning of the slope coefficient β_4 is standard (marginal increase of the sales when income increases by unity accepted) while β_5 shows the increase of the marginal effect in December. The coefficient β_6 could be understood by **two different ways**: on the one hand it is a premium for December for those sales that are made in a big city, one the other hand it is a premium for those sales performed in a big city for the fact of December. **[5 marks]**
[Total 15 marks for a)].

(c) The simple linear regression

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

is to be estimated.

(c1) What is the difference for the estimation of this regression in three situations:

(1) The explanatory variable X_i is a random variable, measured correctly.

(2) X_i is non-stochastic variable, measured with the random error, so in fact the model $Y_i = \alpha_1 + \alpha_2 Z_i + u_i$, $Z_i = X_i + w_i$ is estimated.

(3) Both Y_i and X_i are a non-stochastic variables, measured with some random errors, so in fact the model $Q_i = \gamma_1 + \gamma_2 Z_i + u_i$, $Q_i = Y_i + v_i$ $Z_i = X_i + w_i$ is estimated.

Explain.

(c2) In which of these situations instrumental variable estimation is an improvement on ordinary least squares (OLS). Explain why IV estimation is superior to OLS in this case. Does it mean that IV estimation is more efficient than OLS estimation?

Principles of marking and comments

c1) The solution of this problem reproduces textbook (chapter related to the stochastic explanatory variables). Situation (1) corresponds to the Gauss-Markov theorem (model B), so the estimators are BLUE. Situation (2) is the classic model with the explanatory variables measured with errors, so the reproduction of the proof of inconsistency from the textbook is expected (not presented here). The situation (3) is principally just the same as additional 'noise' in dependent variable caused by the measurement errors is absorbed by the disturbance term of the model what decreases efficiency. **[10 marks]**

c2) Standard proof of consistency of the instrumental variable estimator (according to the textbook) is expected here. As the instrumented variable always has not perfect correlation with the instrument this estimator could be (and in practice usually is) less efficient than conventional OLS estimator, so the potential benefits of using IV estimation could be accompanied by the essential losses in efficiency. **[5 marks]**
[Total 15 marks].

(d) On request of voters the party has published incomes of the candidates. However on those candidates, whose year income X_i exceeded 30 000 dollars the income \$30 000 has been published. Smaller values have not been changed. Actually about 30 percent of all observations have been corrected. The sociologist having data on a sample of n candidates believes that popularity of the candidate (number of the votes given for him/her Y_i) negatively depends on his/her income according to the linear regression

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

where u_i is an independently and identically distributed disturbance term.

The sociologist is trying to estimate this equation using different methods.

- **(d1)** What are consequences of its estimation by OLS using all observations?
- **(d2)** What are consequences of its estimation by OLS using subsample of all not corrected observations?
- **(d3)** What are consequences of its estimation using tobit model? What is the tobit model?

Principles of marking and comments

The answer below assumes $\beta_2 < 0$. There are many ways to approach answering this question. Of course this is something like censoring the data but the censored variable is not dependent but explanatory. So we cannot discuss the typical consequences of censoring data without some additional assumptions relating the knowledge of data by voters and sociologist.
The most general way of discussing this problem is to consider it in terms of specification of the regression model. If the sociologist uses corrected data instead of the real data in d1) he/she would get badly specified model with the typical consequences of incorrect specification (bias, inefficiency, invalidity of tests and so on). This conclusion would not change by the fact whether the voters knows the real data or not.
Situation d2) certainly decreases the volume of data used in estimation, not involving the specification issues.
Tobit model has nothing to do with this situation.

It is also possible to consider this situation in the other way, taking into account that dependent variable Y_i will be also influenced by the cencoring of the explanatory variable. If the researcher uses the distorted data on Y_i but at the same time has real data on explanatory variable it would be situation similar to the censoring of the dependent variable (with the reservation that it is not real censoring). In this situation we could apply (with reservations) the conclusions connected with the censoring: inconsistency and bias **upward** (as $\beta_2 < 0$) in d1, inefficiency and inconsistency in d2, and significant benefits of the using tobit model which combines OLS for not censored data and probit model for the censored data.

There is also the third situation that was discussed in materials published at **mief.hse.ru** for the seminar 14. This discussion is based on the **explicit assumption** that **the values of *Y* have not been altered while X's were censored.** The minor technical difference is that now $\beta_2 < 0$.
We reproduce here this analysis making the necessary changes.
d1). The slope coefficient will tend to be underestimated, the intercept overestimated, and the standard errors invalidated to some extent. Write $Y_i = \beta_1 + \beta_2 Z_i + v_i$ where $Z_i = X_i$ and $v_i = u_i$ for unconstrained observations, and $Z_i = X^{max}$ and $v_i = u_i + \beta_2\left(X_i - X^{\max}\right)$ for constrained observations. The OLS slope coefficient may be decomposed as

$$b_2 = \beta_2 + \sum a_i v_i \quad \text{where} \quad a_i = \frac{Z_i - \bar{Z}}{\sum\left(Z_j - \bar{Z}\right)^2}.$$

For the unconstrained observations, $E(v_i) = 0$. For the constrained observations, $E(v_i) < 0$ and $a_i > 0$. $E(b_2) = \beta_2 + \sum a_i \beta_2 (X_i - X_{\max})$. Since $\left(X_i - X^{\max}\right)$ will be positively correlated with a_i, the direction of the bias will depend on the sign of β_2 (in this situation $\beta_2 < 0$). The standard error will be invalidated by the non-standard distribution of the disturbance term.
d2).There will be no violation of the regression model assumptions and so the parameter estimators will remain unbiased and the standard errors will be valid. However, the variances of the parameter estimators will be larger than those that would have been obtained with the unmodified data because the variance of *X* will be smaller and because the number of observations will be smaller.

[Total 15 marks]

Comments: *almost all student paid no attention to the fact that not dependent but independent variable is censored here and tried to discuss some different situation.*

[Total 60 marks].

2. A researcher is investigating macroeconomic data on the money supply M (currency plus demand deposits), gross domestic product (GDP) – Y, gross private domestic investments I, and government purchases of goods and services G, all in billions of euro for some developed European country from 1992 to 2009, where M and Y are considered as endogenous variables while I and G are exogenous variables. (*Ignore problems connected with non-stationary time series*)

(a) The researcher starts from estimation of two linear equations using OLS (here and hereafter standard errors in parentheses).

$$\hat{M}_t = \underset{(5.1)}{84.8} + \underset{(0.004)}{0.13}Y_t \quad R^2 = 0.98 \qquad (1)$$

$$\hat{Y}_t = \underset{(68.36)}{-385.21} + \underset{(0.66)}{4.81}M_t + \underset{(0.54)}{2.21}I_t \quad R^2 = 0.99 \qquad (2)$$

Why these equations could not be considered as satisfactory?

<u>Principles of marking and comments</u>

a) The first step in the whole solution here is to rewrite econometric model in general (theoretical) form:

$$M_t = \alpha_1 + \alpha_2 Y_t + u_t \qquad (1.1)$$
$$Y_t = \beta_1 + \beta_2 M_t + \beta_3 I_t + v_t \qquad (2.1)$$

To answer the question one should show that the variable Y_t correlates with disturbance term u_t in equation (1.1) and the variable M_t correlates with the disturbance term v_t in equation (2.1). (it is not enough to indicate on the circular dependence of the variables). **[6 marks]**

(b) Then the researcher estimates the following equations

$$\hat{M}_t = 95.86 + 0.80 I_t \quad R^2 = 0.944 \qquad (3)$$
$$(9.85) \quad (0.048)$$

$$\hat{Y}_t = 75.78 + 6.06 I_t \quad R^2 = 0.97 \qquad (4)$$
$$(53.67) \quad (0.26)$$

and derives from here the estimates $b_2 = \frac{0.80}{6.06} = 0.132$ and $b_1 = 95.86 - 0.1321 \cdot 75.78 = 85.85$ to get the following equation

$$\hat{M}_t = 85.85 + 0.132 Y_t \qquad (5)$$

that is quite similar to the equation (1), but there are some differences anyway:
Explain the meaning of these actions, giving necessary mathematical derivations.

Principles of marking and comments

b) These actions constitute the indirect least squares method. It is expected that corresponding reduced form equations are derived from equations (1.1) and (2.1) and then the coefficients of equation (1.1) are derived from reduced form equations. **[7 marks]**

(c) What equation (5) or (1) is better and why? Indicate properties of estimators of coefficients of both equations (5) and (1). Provide mathematical justification for your conclusion.

Principles of marking and comments

c) Any ILS estimator ensures only consistency but does not provides with efficiency it would be of value to consider OLS estimator as reasonable alternative to ILS estimator. Besides the general discussion of comparative advantages and disadvantages of the two estimators it is expected that consistency of some of ILS estimators would be shown. **[7 marks]**

(d) Is it possible to apply the method described in (b) to get alternative estimators for the second equation specified as in (2)? Explain.

Principles of marking and comments

d) No, this is impossible as the reduced form equations provide with less number of algebraic equations than the number of estimated parameters. Alternatively one could prove that the second equation is underidentified. **[6 marks]**

(e) The researcher decided to do the following: he includes additionally an exogenous variable G into equation for Y (which contains already exogenous variable I) and estimates equation

$$\hat{Y}_t = -27.17 + 1.85 I_t + 3.48 G_t \quad R^2 = 0.998 \qquad (6)$$
$$(15.78) \quad (0.29 \quad (0.24)$$

and then substitutes the computed values of $\hat{Y}_t$ into equation for M_t:

$$\hat{M}_t = \underset{(4.50)}{84.4} + \underset{(0.0034)}{0.133}\hat{Y}_t \quad R^2 = 0.989 \tag{7}$$

What is the meaning of this procedure? Compare equation (7) with equation (1). What are advantages and disadvantages of both approaches to the estimation? Explain.

Principles of marking and comments

e) The researcher uses here TSLS method. It is expected that all stages of TSLS are described properly (standard). Providing with the best possible instrument TSLS does not guarantee efficiency of the estimator (as all other IV estimators), so sometimes inconsistent OLS estimators could be a good choice, especially here where all methods give almost similar results (as one could see from comparison of the estimated equations). **[7 marks]**

(f) Would you recommend the researcher to use Durbin-Wu-Haussmann test and why? Explain to the researcher what really could be tested by this test, and how this test could be performed *(No calculations are expected)*.

Principles of marking and comments

f) The similarity of the results of using different methods casts doubt on the hypothesis of endogeneity of the explanatory variables. That is why Darbin-Wu-Haussmann test could be strongly recommended. The original idea of this test is the comparison of the coefficients of equations estimated using different methods (OLS and IV or TSLS) using chi-square statistics with the number of degrees of freedom equal to the number of explanatory variables. The alternative version of this test is known as a Hausman test in the Davidson-MacKinnon version, when residuals of the first step equation of TSLS methods then included in the regression estimation and the significance of the corresponding coefficient is estimated.**[7 marks]**

[Total 40 marks].

3. The following estimates were calculated from a sample of 7,634 women respondents from the General Household Survey 1995. The dependent variable takes the value 1 if the woman was in paid employment and 0 otherwise.

	OLS	**Probit**
HIGH	0.093	0.259
	(0.015)	(0.043)
NOQUAL	– 0.210	– 0.554
	(0.013)	(0.035)
AGE	0.038	0.107
	(0.003)	(0.008)
*AGE*2	– 0.051	– 0.142
	(0.003)	(0.009)
MAR	0.024	0.063
	(0.009)	(0.035)
Constant	– 0.068	– 1.593
	(0.049)	(0.137)

where *HIGH* is 1 if the respondent has a higher educational qualification, 0 otherwise; *NOQUAL* is 1 if the respondent has no professional qualifications, 0 otherwise; *AGE* is age in years; *AGE*2 is

$AGE^2/100$; MAR is 1 if married, 0 otherwise. Conventionally calculated standard errors are in brackets for the ordinary least squares (OLS) results, asymptotic standard errors are in brackets elsewhere.

(a) What are comparative advantages and disadvantages of two models used in estimation of the regression?

Principles of marking and comments

a) Standard. 4 disadvantages of the linear probability model and what are advantages of the logit and probit models. **[6 marks]**

(b) Explain how Probit estimates are calculated on the basis of Maximum Likelihood method.

For the probit model, $f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}Z^2}$ *(the standard normal probability density function).*

Using two sets of estimates, test the null hypothesis that the coefficient of *MAR* is zero. Which test statistics would you consider more reliable? Explain.

Principles of marking and comments

b) It is expected that log likelihood function would be presented. Some general description if the basics of MLE methods are also welcomed.

The coefficient based on MLE estimation is here insignificant, but anyway this method is preferable here as the results of the OLS estimators could be invalid **[6 marks]**

(c) Using OLS and Probit estimates, calculate the estimated probabilities of being in employment for an unmarried woman aged 40 with the higher educational and professional qualification. How these probabilities would change if her age is now 45 keeping other characteristics constant? Comment on your results. *(Instructions: use direct estimation of probabilities in both questions, estimate resulting figures to within 0.01).*

Principles of marking and comments

c)

LPM(40)
=-0.068+0.093*1-0.210*0+0.038*40-0.051*40^2/100
=0.729

LPM(45)
=-0.068+0.093*1-0.210*0+0.038*45-0.051*45^2/100
=0.702

Delta=
=0.702-0.729
-0.023

PROBIT(40)
=-1.593+0.259*1-0.554*0+0.107*40-0.142*40^2/100
0.674
=0.75
(using Normal Tables)

PROBIT(45)
=-1.593+0.259*1-0.554*0+0.107*45-0.142*45^2/100
0.6055
=0.73
(using Normal Tables)

Delta=
=0.73-0.75
-0.02
So the results are approximately similar.
[7 marks]

d) Now evaluate the marginal effect of age on the basis of the Probit model using derivatives. Perform your calculations for an unmarried woman with the higher educational and professional qualification for different ages: 25 years, 40 years, and 55 years. Comment on your results.

Principles of marking and comments
d)
PROBIT(25)
=-1.593+0.259*1-0.554*0+0.107*25-0.142*25^2/100
0.4535
=0.359
(using Normal Tables)
Marginal effect
=0.359*(0.107-2*0.142*25/100)
0.0129

PROBIT(40)
=-1.593+0.259*1-0.554*0+0.107*40-0.142*40^2/100
0.674
=0.317
(using Normal Tables)
Marginal effect
=0.317*(0.107-2*0.142*40/100)
-0.002

PROBIT(55)
=-1.593+0.259*1-0.554*0+0.107*55-0.142*55^2/100
0.255
=0.386
(using Normal Tables)
Marginal effect
=0.386*(0.107-2*0.142*55/100)
-0.019

Conclusion: for young women additional year rises probability of being employed.
For 40 years old woman additional year slightly lowers the probability of being employed.
At the age of 50 the probability of employment drops by 2 percentage point with each year.
All these conclusions are in full correspondence with common sense and econimic reality.
[7 marks]

(e) Evaluate the marginal effect of marriage for an unmarried woman aged 20 having neither educational nor professional qualification using OLS, and Probit estimates. Whether the sign of the estimated effect coincides with your expectations? Comment.

Principles of marking and comments

e)

Directly

Unmarried

LPM

=-0.068+0.093*0-0.210*1+0.038*20-0.051*20^2/100+0.024*0

=0.278

Married

=-0.068+0.093*0-0.210*1+0.038*20-0.051*20^2/100+0.024*1

=0.302

=0.302-0.278

0.024

Using marginal effect

Unmarried

=-1.593+0.259*0-0.554*1+0.107*20-0.142*20^2/100+0.063

=-0.512

P(z)=0.35

Marginal effect

=0.35*0.063

0.022

Almost the same characteristics

The marriage rises the probability of being emploied. There could be many different possible explanation of this fact (really interesting because there are many obvious arguments in favour of the opposite effect). The most interesting fact is that probably men value in women the same qualities as the employers (this similarity of the marriage and emploiment was indicated in many econometrics papers).

[7 marks]

(f) Test the null hypothesis that all the slope coefficients of the Probit model are jointly equal to zero. It is given that

$$\ln L_U = -321.25$$
$$\ln L_R = -416.01$$

where $\ln L_U$ is the log of the likelihood from the unrestricted Probit model (as it presented in the table above) and $\ln L_R$ is the log of the likelihood from the restricted Probit model (the same model estimated with no explanatory variable included in the regression).

Principles of marking and comments

f)

=2*(-321.25+416.01)

=189.52

What is much greater than 1% critical value of chi-square statistics (df=5) equal to 15.086.

[7 marks]

[Total 40 marks].

Comments: *only few students were able to make necessary calculations without errors, and to get correct conclusions.*

The International College of Economics and Finance
Econometrics – 2012-2013.
Mid-year exam. December 24.

Part 2. (Reading time 20 minutes, working time 1 hour 30 minutes).
Answer the first (obligatory) problem and any one of the two problems (2-3).

1. ***All parts a), b), c) d) of this question bear identical number of marks.***

(1a) A researcher has the data of the revenues of two branches of a corporation for 12 months (X_{2t} and X_{3t}) and is interested in how they impact on corporate profit Y_t. Let a regression equation be:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t;\ t = 1, 2, ..., 12$$

Issues related to the non-stationary time series should be ignored. Estimation of this equation gives (standard errors in parentheses):

$$\hat{Y}_t = -16.51 + 1.39 X_{2t} + 0.80 X_{3t};\ t = 1, 2, ..., 12 \quad R^2 = 0.473$$
$$(20.21)\ (0.70) \qquad (0.60)$$

(i) Outline briefly how should she test: - β_2 is different from zero;
- β_3 is different from zero;
- jointly β_2 and β_3 are different from zero;

i) Standard *t*-tests - for β_2: $t = \frac{1.39}{0.70} = 1.99$, for β_3: $t = \frac{0.80}{0.60} = 1.33$, while $t_{crit}^{5\%}(12-3) = 2.262$, so both coefficients are insignificant. For β_2 and β_3 simultaneously: $F = \frac{R^2/(k-1)}{(1-R^2)/(T-k)} = \frac{0.473}{1-0.473} \cdot \frac{9}{2} = 4.04$ ($T = 12$ is a sample size and $k = 3$ is the number of parameters of the model) while $F_{crit}^{5\%}(2,9) = 4.26$, so there is not enough data to establish significant relationship.**[3 marks]**

(ii) How the results in (i) would change if the researcher suggests that X_{2t} and X_{3t} could not influence negatively on the profit Y_t. Perform the appropriate tests and explain the results.

ii) In this case we are allowed to use one-sided test so now $t_{crit}^{5\%}(9) = 1.833$ and $t_{crit}^{1\%}(9) = 2.821$, so the coefficient β_2 becomes significant (only at 5%) and β_3 remains insignificant. No changes for the *F*-test.**[2 marks]**

(iii) The researcher noticed that in the estimated equation the sum of two slope coefficients is approximately equal to 2. How would she test the hypothesis $\beta_2 + \beta_3 = 2$?

iii) Let us redefine the equation $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$ (with $R^2_{unrestr.}$) as follows $Y_t = \beta_1 + (\beta_2 + \beta_3 - 2) X_{2t} - \beta_3 X_{2t} + \beta_3 X_{3t} + 2 \cdot X_{2t} + u_t = \beta_1 + (\beta_2 + \beta_3 - 2) X_{2t} + \beta_3 (X_{3t} - X_{2t}) + 2 X_{2t} + u_t$
Running regression $Y - 2X_{2tt} = \beta_1 + (\beta_2 + \beta_3 - 2) X_{2t} + \beta_3 (X_{3t} - X_{2t}) + u_t$ and finding $R^2_{restr.}$ one can use either two-sided *t*-test for coefficient $\beta_2 + \beta_3 - 2$ or an *F*-test for restriction $\frac{R^2_{unrestr.} - R^2_{restr.}}{(1 - R^2_{unrestr.})/(T-k)}$
.**[2 marks]**

(iv) From the description above it follows that coefficients $\hat{\beta}_2$ and $\hat{\beta}_3$ are comparable by their meaning. The researcher found that in the estimated equation $\hat{\beta}_2 > \hat{\beta}_3$. How to test this idea in relation to the theoretical coefficients (state null and alternative hypotheses and suggest a procedure of testing)?

iv) Now we also redefine equation $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$ as follows $Y_t = \beta_1 + (\beta_2 - \beta_3) X_{2t} + \beta_3 X_{2t} + \beta_3 X_{3t} + u_t = \beta_1 + (\beta_2 - \beta_3) X_{2t} + \beta_3 (X_{3t} + X_{2t}) + u_t$ but now we should use one-sided t-test for the coefficient $(\beta_2 - \beta_3)$.**[3 marks]**

(1b) A researcher is studying interaction of printed publications and online media activity of several publishing houses $k = 1, 2, ..., n$. She has data on OM_k - rate of growth of income from online media (Internet publications) in the current year, PP_k - rate of growth of income from printed publications, I_k - total investments, and RD_k - the share of research and development of new media in expenditures of publishing house. The variables OM_k and PP_k are supposed to be endogenous, while I_k and RD_k are exogenous. The researcher is considering the following model:

$$\begin{aligned} PP_k &= \alpha_1 + \alpha_2 OM_k + \alpha_3 I_k + u_{1k} \\ OM_k &= \beta_1 + \beta_2 PP_k + \beta_3 RD_k + u_{2k} \end{aligned} ; \ k = 1, 2, ..., n.$$

(i) Examine the identification of both equations. How would you estimate each of the equations? If there are several options for choosing the method of estimation give some reasons pro and contra each method suggested.

i) Both equations are exactly identified as in each equations the number of missed variables is equal to the number of equations in a system. As there is a circular dependence between endogenous variables the disturbance terms of both equations are correlated with some regressors, so Gauss-Markov conditions are violated and so OLS estimators of coefficients are biased and inconsistent. To estimate both equations ILS, IV or TSLS methods could be used. TSLS gives here the same results as IV as both equations are exactly identified.**[5 marks]**

(ii) A colleague of the researcher commented that investments I_k should be also included into second equaiton for online media. How this would change the identification and estimation of both equations?

ii) If I_k is included into second equation it becomes underidentified as there is no instrument for endogenous variable PP_k. First equation remains exactly identified..**[3 marks]**

(iii) At the seminar one of the participants gave the researcher a fresh idea to use last year data on growth rate of online media $OM_k(-1)$ in the system of simultaneous equations. But he did not specify the way how this wariable should be included. Suppose that a proposal of the colleague mentioned in (ii) has taken into account so the variable I_k is already included into second equaiton.
Discuss different approaches to inclusion of the variable $OM_k(-1)$ into one or two of the equations and compare the consequences of different approaches for the estimation of both equations.

iii) The variable $OM_k(-1)$ is fully determined in the last period so it could be considered as exogenous vaiable (so called predetermined variable). There are three options of including it into the system. If it is included into first equation it could be used as an instrument for endogenous variable PP_k and so both equations become exactly identified. If it is included into second equation it remains underidentified and the first equation becomes overidentified (now there are two instruments RD_k and $OM_k(-1)$ for endogenous variable OM_k). Inclusion of $OM_k(-1)$ into both equations changes nothing. Also $OM_k(-1)$ could be simply used as additional instument for TSLS method. **[3 marks]**

1c) Working on her coursework a student of ICEF collected data on a sample of 247 ICEF graduates from different years of graduation working in Russia. She is interested in studying their current salary – ***SAL*** (in US thousands of dollars per year). Explanatory variables are ***AGE*** (age of respondent), age squared – ***AGE2***

($AGE2 = AGE^2$), and also some dummy variables: ***MSCA*** (Master of Science Degree Abroad – it is equal to 1 for those graduates who have received a master's degree abroad, and 0 otherwise), ***NFE*** (no further education - equal 1 for those graduates who received master's degree neither abroad nor in the country), and ***MALE*** (equal 1 for male and 0 for female). Here you have the results of estimation of 2 regressions using different sets of variables (standard errors in brackets).

$$\hat{SAL}_i = \underset{(0.18)}{15.18} + \underset{(0.021)}{0.254}AGE_i - \underset{(0.00005)}{0.001}AGE2_i \qquad R^2 = 0.48 \qquad (1)$$

$$\hat{SAL}_i = \underset{(0.049)}{10.068} + \underset{(0.03)}{0.18}AGE_i - \underset{(0.00003)}{0.0008}AGE2_i + \underset{(0.456)}{4.021}MSCA_i - \underset{(2.237)}{5.123}NFE_i + \underset{(0.016)}{0.025}MALE_i \qquad R^2 = 0.63 \qquad (2)$$

(i) How many categories of education level of ICEF graduates describe dummy variables ***MSCA*** and ***NFE***? What is the reference category for the equation (2)? Why coefficients of common variables are different in equations (1) and (2)?

i) Three categories: no further education, master's degree in Russia and master's degree abroad. Equation (2) contains also dummy variable ***MALE*** ; the reference category corresponds to the values of all dummy variables equal to zero, so it is female person getting master's degree in Russia. Equation (2) allows describe the salary function for different categories of ICEF graduates (males and females, getting master's degree or not etc., while equation (1) describes salary function of all ICEF garaduates.**[2 marks]**

(ii) What is the difference between interpretations of equations (1) and (2) (give a formal interpretations not paying attention to the significance of coefficients and common sense).

ii) In equation (1) the constant term formally gives the 'initial' salary for the ICEF graduate of zero years old. As it has no sense it would be better to use in interpretation average age of ICEF graduates for the moment of graduation $\overline{AGE}$, then $15.18 + 0.254\overline{AGE} - 0.001\overline{AGE}^2$ shows the starting salary of a 'fresh' ICEF graduate. The similar expression $10.068 + 0.18\overline{AGE} - 0.0008\overline{AGE}^2$ gives starting salary of a 'fresh' female ICEF graduate having intention to get master's degree in Russia. Annual increase in salary of a graduate of the age AGE_i according equation (1) is $\frac{\partial SAL}{\partial AGE} = 0.254 - 2 \cdot 0.001 AGE_i = 0.254 - 0.002 AGE_i$, while according equation (2) the annual increase is $0.18 - 0.0016\overline{AGE}_i$. Two estimates of marginal effects of ***AGE*** are different as equation (1) lacks many important variables. The interpretation of other coefficients is standard as marginal effects of corresponding variables on salary. **[2 marks]**

(iii) Are the coefficients of the variables ***MSCA***, ***NFE*** and ***MALE*** significant? Are they jointly significant? What they tell about salaries of ICEF graduates?

iii) Absolute values of *t*-statistics for these variables are 8.82, 2.29, 1.56 while $2.576 = t_{crit}^{1\%}(\infty) < t_{crit}^{1\%}(241) < t_{crit}^{1\%}(120) = 2.617$, $1.960 = t_{crit}^{5\%}(\infty) < t_{crit}^{5\%}(241) < t_{crit}^{5\%}(120) = 1.980$ so the coefficient of ***MSCA*** is significant at 1%, the one of ***NFE*** is significant only at 5% while the coefficient of ***MALE*** is insignificant. Finding *F*-statistics $F = \frac{(R_u^2 - R_r^2)/3}{(1 - R_u^2)/241} = \frac{(0.63 - 0.48)/3}{(1 - 0.63)/241} = 32.57$ and comparing it with $F_{crit}^{1\%}(3, 241) < 3.95$ we conclude that these variables as a group are significant. **[2 marks]**

(iv) What is the reason to include variable ***AGE2*** into equations? What additional information gives this variable and how its presence changes the coefficients of the variable ***AGE***? How the coefficients of the model would change, if the variable ***AGE2*** would be eliminated from the equations.

iv) ***AGE2*** allows to get into account the decreasing marginal productivity of ***AGE*** **[1 mark]**. If the variable ***AGE2*** is omitted from equation the estimate of the coefficient β_{AGE} would be biased, the bias is $\beta_{AGE2}\dfrac{\mathrm{Cov}(AGE, AGE2)}{\mathrm{Var}(AGE)}$. An one could expect $\beta_{AGE2} < 0$ and $\mathrm{Cov}(AGE, AGE2)$ is obvious;y positive, the bias is negative so $\hat{\beta}_{AGE}$ decreases. **[1 mark]**

(v) At the seminar where results of this study were discussed the participants suggested to include two additional variables into equation (2):

- one of them suggested to take into account the variable ***RATING*** (overall rating of a graduate during her/his study at ICEF in %);
- the other suggested to take into account also variable ***EXP*** (work experience in years).

Discuss each of these suggestions: are they reasonable? What would have been the consequences for the estimation of regression if each of these proposal was accepted?

v) ***RATING*** is quite a good proxy for the expectation of the former ICEF student's productivity that obviously lacks in the set of available data **[1 mark]**. ***EXP*** would be useful but it highly correlates with ***AGE*** and its inclusion could lead to the danger of multicollinearity. **[1 mark]**

(1d) Let $x_1, x_2, \ldots, x_n$ be a random sample from a distribution with the probability density function of

$$f(x) = \alpha x^{\beta}, \qquad 0 \le x \le 1.$$

(i) What is relationship between α and β so that $f(x) = \alpha x^{\beta}$, $0 \le x \le 1$ is really probability density function of a random variable?

i) A function $f(x) = \alpha x^{\beta}$, $\quad 0 \le x \le 1$ is a density distribution function if $\int_{-\infty}^{+\infty} f(x)dx = 1$, so $\int_0^{+1} \alpha x^{\beta} dx = \dfrac{\alpha}{\beta+1} x^{\beta+1}\Big|_0^{+1} = \dfrac{\alpha}{\beta+1} = 1$, so $\beta = \alpha - 1$, and $f(x) = \alpha x^{\alpha-1}$, $\quad 0 \le x \le 1$. **[2 marks]**

(ii) Explain, what are MLE estimators of distribution parameters.

ii) MLE estimators of distribution parameters are those that maximize likelyhood function, or the probability to get the observed sample: $P(x_1; x_2; \ldots; x_n \mid \alpha, \beta) \xrightarrow[\alpha,\beta]{} \max$. **[2 marks]**

(iii) Derive MLE estimators of parameters α and β on the basis of sample $x_1, x_2, \ldots, x_n$

iii) Likelihood function is $L = L(x_1; x_2; \ldots; x_n \mid \alpha) = \alpha x_1^{\alpha-1} \cdot \alpha x_2^{\alpha-1} \cdot \ldots \alpha x_n^{\alpha-1} = \alpha^n (x_1 \cdot x_2 \ldots x_n)^{\alpha-1}$, so Log likelihood function is $\ln L = n \ln \alpha + (\alpha - 1)\ln(x_1 \cdot x_2 \ldots x_n)$. Using first order condition we get $\dfrac{d \ln L}{d\alpha} = \dfrac{n}{\alpha} + \ln(x_1 \cdot x_2 \ldots \cdot x_n) = 0$, so $\alpha = -\dfrac{n}{\ln(x_1 \cdot x_2 \ldots \cdot x_n)}$ or $\alpha = -\dfrac{n}{\sum \ln(x_i)}$. **[4 marks]**

(iv) Let the sample be $0.1,\ 0.3,\ 0.5,\ 0.7,\ 0.9$. Evaluate the estimators of parameters obtained in (i–ii) and investigate whether they give a reasonable description of the sample distribution.

iv) For the sample under consideration we get $\alpha = -\dfrac{5}{\ln(0.1 \cdot 0.3 \cdot 0.5 \cdot 0.7 \cdot 0.9)} \approx 1.07$, so $f(x) = 1.07 \cdot x^{0.07} \approx 1$ that is reasonable for the nearly uniform distribution of the sample values on the interval $0 \le x \le 1$. **[2 marks]**

[Total 40 marks for Question 1 + possible bonus marks)].

2. A researcher is studying the factors affecting the AP exams results in macro and microeconomics of 178 ICEF first year students. Based on data for the 2010-2011 school year, the researcher considers a system of simultaneous equations relating the results of the AP examinations in macro (***MAC***) and microeconomics (***MIC***) with the results of previously passed State math exam (***EGEM***), State English exam (***EGEE***), as well as the gender of the students (***MALE***) and the fact of enrollment at ICEF on the basis of the results of one of the Olympiads (***OL***):

$$MAC_i = \alpha_1 + \alpha_2 MIC_i + \alpha_3 EGEM_i + \alpha_4 MALE_i + \alpha_5 OL_i + u_i$$
$$MIC_i = \beta_1 + \beta_2 MAC_i + \beta_3 EGEM_i + \beta_4 MALE_i + \beta_5 OL_i + \beta_6 EGEE_i + v_i$$

Using OLS she estimates these equations (here and further standard errors in parentheses)

$$\hat{MAC}_i = \underset{(0.42)}{0.66} + \underset{(0.05)}{0.97}MIC_i - \underset{(0.006)}{0.002}EGEM_i + \underset{(0.11)}{0.23}MALE_i - \underset{(0.17)}{0.15}OL_i \quad R^2 = 0.72 \tag{1}$$

$$\hat{MIC}_t = \underset{(0.70)}{-1.97} + \underset{(0.04)}{.72}MAC + \underset{(0.006)}{0.025}EGEM - \underset{(0.10)}{0.08}MALE + \underset{(0.15)}{0.32}OL + \underset{(0.007)}{0.01}EGEE_i \quad R^2 = 0.75 \tag{2}$$

(2a) Comment on the structure of significant relationships between variables under stated assumptions (no need to describe significance tests in details). Comment on the identification of each equation supposing variables ***MAC*** and ***MIC*** being endogenous and variables ***EGEM***, ***EGEE***, ***MALE***, and ***OL*** being exogenous.

Principles of marking and comments

a) In both equations both endogenous variables are highly significant. In equation for *MAC* variable *MALE* is also significant at 5% level. In the equation for *MIC* variable *MALE* is insignificant, but variables *EGEM* and *OL* are significant. One should notice that the difference in structure of significant coefficients could be explained not only by the real impact of explanatory variables but also by the close interaction of endogenous variables. **[2 marks]**

First equation (for *MAC*) is identified as one variable is missed from it. The second equation is underidentified. The second equation is underidentified as no variable is missed from it. **[3 marks]**
[Total 5 marks for a)]

(2b) Why equations (1-2) cannot be considered as satisfactory? What is indirect least squares (ILS) method? What are properties of ILS estimators? Is it possible to get estimates of α_2 and β_2 in the original equations using ILS method?
ILS (Indirect Least Squares) method consists of estimation of reduced form equations and then some attempts to derive from them the estimators for original structural equations.**[1 mark]**

Reduced form system of equations is here:

$$MAC_i = \frac{1}{1-\alpha_2\beta_2}[(\alpha_1+\alpha_2\beta_1)+(\alpha_3+\alpha_2\beta_3)EGEM_i+(\alpha_4+\alpha_2\beta_4)MALE_i+(\alpha_5+\alpha_2\beta_5)OL_i+\alpha_2\beta_6 EGEE_i+w_{1i}$$

$$MIC_i = \frac{1}{1-\alpha_2\beta_2}[(\beta_1+\alpha_2\beta_2)+(\beta_3+\alpha_3\beta_2)EGEM_i+(\beta_4+\alpha_4\beta_2)MALE_i+(\beta_5+\alpha_5\beta_2)OL_i+\beta_6 EGEE_i+w_{2i}$$

where $w_{1i} = u_i + \alpha_2 v_i$, $w_{2i} = v_i + \beta_2 u_i$ **[2 marks]**

Both reduced form equations contain only exogenous variable in their right-hand sides so OLS applied to them gives BLUE estimates of coefficients.

$$\hat{MAC}_i = \pi_1 + \pi_2 EGEM_i + \pi_3 MALE_i + \pi_4 OL_i + \pi_5 EGEE$$
$$\hat{MIC}_i = \tau_1 + \tau_2 EGEM_i + \tau_3 MALE_i + \tau_4 OL_i + \tau_5 EGEE$$
[1 marks]

There are many relationships between coefficients of structural and reduced form equations, by example

$$\frac{\alpha_2\beta_6}{1-\alpha_2\beta_2} = \pi_5;$$

$$\frac{\beta_6}{1-\alpha_2\beta_2} = \tau_5$$

So dividing first of the last two equations by the second one we get $\alpha_2 = \frac{\pi_5}{\tau_5}$. **[2 marks]**

This ILS estimator is consistent. **[1 mark]** (bonus 2 marks for the proof)

There is likely no way to get estimator for β_2 as second equation is underidentified. **[1 mark]**

[Total 8 marks for b)]

(2c) The researcher is interested most in the estimation of coefficients α_2 and β_2 of the endogenous variables ***MIC*** and ***MAC***, describing their interaction. Show that OLS estimators of these coefficients are biased and discuss the direction of the bias.

Omitting variables $EGEM_i$, $MALE_i$ and OL_i (otherwise we need matrix algebra) and so reducing structural equations to

$$MAC_i = \alpha_1 + \alpha_2 MIC_i + u_i$$
$$MIC_i = \beta_1 + \beta_2 MAC_i + \beta_6 EGEE_i + v_i$$

we will show that OLS estimator of α_2 is biased.

The corresponding reduced form equations are

$$MAC_i = \frac{1}{1-\alpha_2\beta_2}[(\alpha_1 + \alpha_2\beta_1) + \alpha_2\beta_6 EGEE_i + u_i + \alpha_2 v_i]$$
$$MIC_i = \frac{1}{1-\alpha_2\beta_2}[(\beta_1 + \alpha_2\beta_2) + \beta_6 EGEE_i + v_i + \beta_2 u_i]$$

[1 mark] for starting point and simplification of the system – it is impossible to do anything without this!

Remark: To make the structure of the answer more clear it is possible here and thereafter to rename the variables as $MAC_i = Y_{1i}$, $MIC_i = Y_{2i}$, $EGEM_i = X_{1i}$, $MALE_i = X_{2i}$, $OL_i = X_{3i}$, $EGEE_i = X_{4i}$, so the structural equations be

$$Y_{1i} = \alpha_1 + \alpha_2 Y_{2i} + \alpha_3 X_{1i} + \alpha_4 X_{2i} + \alpha_5 X_{3i} + u_i$$
$$Y_{2i} = \beta_1 + \beta_2 Y_{1i} + \beta_3 X_{1i} + \beta_4 X_{2i} + \beta_5 X_{3i} + \beta_6 X_{4i} + v_i$$

and so on.

Using Covariance Rules

$$\hat{\alpha}_2^{OLS} = \frac{\text{Cov}(MAC, MIC)}{\text{Var}(MIC)} = \frac{\text{Cov}([\alpha_1 + \alpha_2 MIC + u], MIC)}{\text{Var}(MIC)}$$
$$= \frac{\text{Cov}(\alpha_1, MIC) + \text{Cov}(\alpha_2 MIC, MIC) + \text{Cov}(u, MIC)}{\text{Var}(MIC)} = \alpha_2 + \frac{\text{Cov}(u, MIC)}{\text{Var}(MIC)}.$$ **[1 mark]**

(Cov(γ, G) is 0 because γ is a constant).

Of course, we would like it to have expected value 0, making the estimator unbiased. However, the error term is a nonlinear function of both *u* and *v* because both are components of *MIC*. As a consequence, it is not possible to obtain an analytical expression for its expected value.

We will investigate the large-sample properties instead. We will demonstrate that the estimator is inconsistent, and this will imply that it has undesirable small-sample properties.

We will start by deriving the limiting value of the numerator, substituting for *MIC* from its reduced form equation.

$$\text{plim Cov}(u, MIC) = \text{plim Cov}\left(u, \frac{1}{1-\alpha_2\beta_2}[(\beta_1+\alpha_2\beta_2)+\beta_6 EGEE_i + v_i + \beta_2 u_i]\right) =$$

$$= \frac{1}{1-\alpha_2\beta_2}[\text{plim Cov}(u,(\beta_1+\alpha_2\beta_2) + \text{plim Cov}(u, \beta_6 EGEE) + \text{plim Cov}(u, v) + \text{plim Cov}(u, \beta_2 u)] = \frac{\beta_6\sigma_u^2}{1-\alpha_2\beta_2}$$

[2 marks]

We use Covariance Rule 1 to decompose the expression.
The first term is 0 because $\beta_1+\alpha_2\beta_2$ is a constant. The second term is 0 because $EGEE$ is exogenous and so distributed independently of v. The third term is 0 under assumption that the disturbance terms are distributed independently of each other. However, the last term is nonzero because the limiting value of a sample variance is the corresponding population variance.

[2 marks]

Next we will derive the limiting value of Var(*MIC*). Variances are unaffected by additive constants, so the first part of the expression may be dropped.

$$\begin{aligned}
\text{plim Var}(MIC) &= \text{plim Var}\left(\frac{\beta_1+\alpha_2\beta_2}{1-\alpha_2\beta_2} + \frac{\beta_6 EGEE + v + \beta_2 u}{1-\alpha_2\beta_2}\right) \\
&= \text{plim Var}\left(\frac{\beta_6 EGEE + v + \beta_2 u}{1-\alpha_2\beta_2}\right) \\
&= \left\{\frac{1}{(1-\alpha_2\beta_2)^2}\right\}\left\{\begin{aligned}&\text{plim Var}(\beta_6 EGEE) + \text{plim Var}(v) \\ &+ \text{plim Var}(\beta_2 v) + 2\text{plim Cov}(\beta_6 EGEE, v) \\ &+ 2\text{plim Cov}(\beta_6 EGEE, \beta_2 u) \\ &+ 2\text{plim Cov}(v, \beta_2 u)\end{aligned}\right\}
\end{aligned}$$

The limiting values of the variances are the corresponding population variances. The limiting values of the covariances are all 0 on the assumption that *EGEE*, *u*, and *v* are distributed independently of each other.

[3 marks]

Hence we obtain an expression for the limiting value of the OLS estimator of the slope coefficient. We can see that the estimator is inconsistent

$$\hat{\alpha}_2^{OLS} = \alpha_2 + \frac{\text{Cov}(u, MIC)}{\text{Var}(MIC)}$$

$$\text{plim Cov}(v, G) = \frac{\beta_6\sigma_u^2}{1-\alpha_2\beta_2}$$

$$\text{plim Var}(MIC) = \left\{\frac{\beta_6^2\sigma_{EGEE}^2 + \sigma_v^2 + \beta_2^2\sigma_u^2}{(1-\alpha_2\beta_2)^2}\right\}$$

$$\text{plim}\ \hat{\alpha}_2^{OLS} = \alpha_2 + (1-\alpha_2\beta_2)\frac{\beta_6\sigma_u^2}{\beta_6^2\sigma_{EGEE}^2 + \sigma_v^2 + \beta_2^2\sigma_u^2}$$

Can we determine the sign of the bias?
The sign of the bias will depend on the sign of the coefficient β_6 and the sign of the term $1-\alpha_2\beta_2$. Since all the variance components are positive the bias is probably positive.

Sign of β_6 is probably positive from common sense considerations – better knowledge of English assessed by *EGEE* generally increases results in Microeconomics) (by looking at the reduced form equations, we can argue that this term should be positive; *MIC* and *MAC* should both be increasing functions of *EGEE*. β_2, α_2 and β_6 should all be positive; so $1-\alpha_2\beta_2$ must also be positive).

[1-2 bonus marks] for the discussion of the signs of the expressions.
[Total 10 marks for c)].

(2d) To get consistent estimators of α_2 and β_2 the researcher first uses instrumental variable ***CALC*** (AP Calculus exam result) and dropping some explanatory variables obtains

$$\hat{\alpha}_2 = \frac{\mathrm{Cov}(MAC, CALC)}{\mathrm{Cov}(MIC, CALC)} = 1.02 \text{ and } \hat{\beta}_2 = \frac{\mathrm{Cov}(MIC, CALC)}{\mathrm{Cov}(MAC, CALC)} = 0.98.$$

Then she estimates the following equation, using additional instrument ***STAT*** (AP Statistics exam result)

$$MIC_i = \gamma_1 + \gamma_2 EGEM_i + \gamma_3 MALE_i + \gamma_4 L_i + \gamma_5 EGEE + \gamma_6 CALC + \gamma_7 STAT + w_i \qquad (3)$$

and uses estimated values $\hat{MIC}_i$ instead of MIC_i in OLS estimation of equation for ***MAC***

$$\underset{}{\hat{MAC}_i} = \underset{(0.51)}{0.95} + \underset{(0.07)}{1.06}(\hat{MIC})_i - \underset{(0.008)}{0.01}EGEM_i + \underset{(0.13)}{0.18}MALE_i - \underset{(0.19)}{0.25}OL_i \quad R^2 = 0.64 \qquad (4)$$

Comment on these actions and their results. What are advantages and disadvantages of different approaches to the estimation? Explain.

First two estimators are instrumental variables estimators. They are also based on the simplification of the structure of the model. These estimators are generally speaking consistent **[1 mark]**

[2 bonus marks] for the proof of the inconsistency)
[2 add.bonus marks if student understand that simplification of the model leads to the omitted variable bias**]**

The second method is a TSLS (Two Stage Least Squares). At the first stage the researcher using all available instruments evaluated estimated values of endogenous variable $\hat{MIC}_i$. At the second stage she uses this new instrument $\hat{MIC}_i$ to get consistent estimates of all coefficients of the first equation for *MAC*. **[2 mark]**
There are two main advantages in this method as to compare to IV method:

1) When there are several available instruments each of the instruments could provide with its own IV estimate, so there is a conflict between them; TSLS method find the BEST (most efficient) combination of all possible instruments, so combined instrument provided with TSLS is generally speaking better than IV. **[2 marks]**
2) There is no omitted variable in the process so there is no additional bias. **[1 bonus mark]**

Using new instrument *CALC* is based on the implicit assumption that *CALC* is a valid instrument (is exogenous, correlated with *MAC* but not correlated with the disturbance term of the equation for *MIC*.
The same should be noted on the variable *STAT* used in TSLS method. **[1 bonus mark]**

It should be noted also that consistency does not imply efficiency so the estimates of IV and TSKS method could have quite big standard errors. We observe that all three method (OLS, IV, TSLS) in our case give quite similar results. . **[1 bonus mark]**

Using new instrument *CALC* is based on the implicit assumption that *CALC* is a valid instrument (is exogenous, correlated with *MAC* but not correlated with the disturbance term of the equation for *MIC*.
The same should be noted on the variable *STAT* used in TSLS method.**[1 bonus mark]**

Adding two exogenous variables to both equation makes them both iveridentified.**[1 bonus mark]**

[Total 5 marks for d].

(2e) Now the researcher uses equation (3) in different way: she memorizes its residuals ***RES*** and substitutes them as an additional variable into original equation

$$\underset{(0.43)\ (0.06)\qquad\quad (0.007)\qquad\quad (0.11)\qquad\quad (0.16)\qquad (0.09)}{\hat{MAC}_i = 0.96 + 1.06MIC_i - 0.01EGEM_i + 0.18MALE_i - 0.25OL_i - 0.45RES \quad R^2 = 0.75} \tag{5}$$

What is the aim of this procedure? What should be the next step in this procedure? What are conclusions from this analysis? Explain.

The researcher performed Hausman test (in the Davidson-MacKinnon form) for the endogeneity of the variable *MIC* **[2 marks]**. After adding memorized residuals into original equation for *MAC* the researcher should test whether coefficient of the variable *RES* is significant. **[1 mark]**

As it is in fact significant $t = \frac{-0.45}{0.09} = 5$ **[2 marks]** the null hypothesis of no endogeneity **[1 marks]** is rejected **[1 marks]**. It means that researcher's fears of endogeneity and consequently of biasedness of OLS estimators were well-founded. **[1 bonus marks]**

[Total 7 marks for e)]

[Total 35 marks for Question 2 + possible bonus marks)].

3. A variable Y is determined by a variable X, the relationship being

$$Y = \beta_1 + \beta_2 X + u$$

where u is a disturbance term that satisfies the regression model assumptions. The values of X are drawn randomly from a population with variance σ_X^2. A researcher makes a mistake and regresses X on Y, fitting the equation

$$\hat{X} = d_1 + d_2 Y$$

where $d_2 = \dfrac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (Y_i - \bar{Y})^2}$. When he realizes his mistake, he points out that the original relationship could be rewritten

$$X = -\frac{\beta_1}{\beta_2} + \frac{1}{\beta_2} Y - \frac{1}{\beta_2} u$$

and hence d_2 will be an estimator of $\dfrac{1}{\beta_2}$. From this he could obtain an estimate of β_2.

(3a) What is the difference in assumptions of so called Model A and Model B? What is in common and what is different in consequences of these assumptions? Explain why it is not possible to derive a closed-form expression for the expected value of d_2 for a finite sample.

The common part of assumptions of Model A and model B include the following:

-The regression model is linear in parameters and correctly specified

-There does not exist an exact linear relationship among the regressors.

-The disturbance term has zero expectation

-The disturbance term is homoscedastic (its variance is constant)

-The values of the disturbance term have independent distributions. **[2 marks** for full list of assumptions, no mark if 3 of them or less mentioned**]**.

(Additional assumption on normality of disturbance term for each its value is used usually to ensure validity of the tests, but it has nothing to do with Gauss-Markov conditions, that ensure that OLS estimators are blue). **[1 bonus mark]**.

There is some differences between two models. In model B the explanatory variables are stochastic **[1 mark]**. It means that the values of regressors are drawn randomly from fixed population **[1 mark]**. So additional assumption is included: the disturbance term is distributed independently of the regressors **[1 mark]**

It can be weakened as follows: The disturbance term has zero conditional expectation **[1 bonus mark]**

Under these assumptions it could be shown that OLS estimators of coefficients for the Model B are unbiased **[1 mark]** **even for the finite samples** **[1 bonus mark]**. These estimators are also consistent **[the same bonus mark as previous,** not additional bonus mark**].** For the model A estimators of coefficients are BLUE as it is well known. [no mark]

Provided that these estimators are regarded as being conditional on the sample values of the regressor(s), the expressions for the variances of the regression coefficients for the simple regression model and the multiple regression model with two regressors remain valid. Likewise the Gauss-Markov theorem remains valid in this conditional sense. **[2 bonus mark]**

One cannot obtain a closed-form expression for the expectation of d_2 because both the numerator and the denominator are functions of u. **[2 mark]**

[Total 8 marks for a)]

(3b) Demonstrate that d_2 is an inconsistent estimator of $\frac{1}{\beta_2}$. Find if it is possible, large sample bias and determine its direction.

$$d_2 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(Y_i - \bar{Y})^2}$$
$$= \frac{\sum(X_i - \bar{X})(\beta_1 + \beta_2 X_i + u_i - \beta_1 - \beta_2 \bar{X} - \bar{u})}{\sum(\beta_1 + \beta_2 X_i + u_i - \beta_1 - \beta_2 \bar{X} - \bar{u})^2}$$
$$= \frac{\frac{1}{n}\sum \beta_2 (X_i - \bar{X})^2 + \frac{1}{n}\sum(X_i - \bar{X})(u_i - \bar{u})}{\frac{1}{n}\sum \beta_2^2 (X_i - \bar{X})^2 + \frac{1}{n} 2\beta_2 \sum(X_i - \bar{X})(u_i - \bar{u}) \frac{1}{n}\sum(u_i - \bar{u})^2}$$

[3 marks]

Hence

$$\text{plim}\, d_2 = \frac{\beta_2 \operatorname{var}(X) + \operatorname{cov}(X,u)}{\beta_2^2 \operatorname{var}(X) + 2\operatorname{cov}(X,u) + \operatorname{var}(u)} = \frac{1}{\beta_2}\left(\frac{1}{1 + \frac{1}{\beta_2^2}\frac{\sigma_u^2}{\sigma_X^2}}\right)$$ **[3 mark]**

since $\operatorname{cov}(X,u) = 0$. **[1 mark]** The estimator will tend to underestimate $1/\beta_2$ in absolute terms **[2 mark]**.
[Total 9 marks for d)]

(3c) Suppose that there exists a third variable Z that is correlated with Y but independent of u. Demonstrate that if the researcher had regressed X on Y using Z as an instrument for Y, the slope coefficient d_2^{IV} would have been a consistent estimator of $\frac{1}{\beta_2}$.

$$d_2^{\text{IV}} = \frac{\sum(Z_i - \bar{Z})(X_i - \bar{X})}{\sum(Z_i - \bar{Z})(Y_i - \bar{Y})} = \frac{\sum(Z_i - \bar{Z})(X_i - \bar{X})}{\sum(Z_i - \bar{Z})(\beta_2 X_i + u_i - \beta_2 \bar{X} - \bar{u})}$$
$$= \frac{\frac{1}{n}\sum(Z_i - \bar{Z})(X_i - \bar{X})}{\beta_2 \frac{1}{n}\sum(Z_i - \bar{Z})(X_i - \bar{X}) + \frac{1}{n}\sum(Z_i - \bar{Z})(u_i - \bar{u})}$$ **[2 marks]**

Hence

$$\text{plim}\, d_2^{\text{IV}} = \frac{\operatorname{cov}(Z,X)}{\beta_2 \operatorname{cov}(Z,X) + \operatorname{cov}(Z,u)} = \frac{1}{\beta_2}$$ **[3 marks]**

since $\operatorname{cov}(Z,u) = 0$ **[1 mark]**.
So d_2^{IV} yields a consistent estimator because, by assumption, Z is independent of u **[1 mark]**.
[Total 7 marks for d)]

(3d) Explain, with reference to the regression model assumptions, why d_2 yielded an inconsistent estimate of $\frac{1}{\beta_2}$ while d_2^{IV} yielded a consistent one.

d_2 is an inconsistent estimator of $1/\beta_2$ because Y and u are correlated (as it was shown in b). d_2^{IV} yields a consistent estimator because, by assumption, Z is independent of u (as it was shown in c). **[3 marks]**
[Total 3 marks for d)]

(3e) What is the difference between instrumental variable (instrument) and proxy variable? What are rules for choosing valid instruments? Is it possible that *X* might be a valid instrument?
During discussion of these results some of the participants suggested to use TSLS approach instead saying that it has many advantages compared to the instrumental variable approach. Comment on this suggestion.

Proxy variables are used instead of missed and unavailable explanatory variables that are important in explanation of fluctuation of the dependent variable. The aim of introducing proxy variable is to prevent omitted variable bias. The instrumental variable is introduced instead of explanatory variable to prevent or mitigate (slack) the violation of Gauss-Markov condition when stochastic explanatory variable is correlated with the disturbance term of the equation. **[2 marks]**

Using instrument is based on the following assumptions: a variable is a valid instrument if it correlates with instrumented explanatory variable but not correlated with disturbance term of the equation for dependent variable.**[1 mark]**

X would be the best possible instrument because, being the determinant of *Y*, it is the variable **most highly correlated** with *Y* **[1 mark]**. It does not appear on the right side of the regression equation and it is independent of *u* so it is valid instrument. (за это уже давали баллы, так то новых не нужно давать)

As there is only one instrumental variable and it is valid instrument both approaches (IV and TSLS) to estimation give identical results **[1 mark]** .
It could be easily proved in general terms. Let the model be $Y_i = \beta_1 + \beta_2 X_i + u_i$ satisfying all assumptions of the Model B except the assumption on X_i being distributed independently of u_i , and let Z be a valid instrument for X. Then $\hat{\beta}_2^{\text{IV}} = \dfrac{\text{Cov}(Y,Z)}{\text{Cov}(Z,X)}$. In TSLS at the first stage X_i is regressed on Z to get estimation of coefficient γ_2 in a regression $X_i = \gamma_1 + \gamma_2 Z_i + v_i$: $\gamma_2^{OLS} = \dfrac{\text{Cov}(X,Z)}{\text{Var}(Z)}$; and to get then estimated values of $\hat{X}_i = \gamma_1 + \gamma_2^{OLS} Z_i$. At the second stage the regression $Y_i = \beta_1 + \beta_2 \hat{X}_i + w_i$ is estimated. The OLS estimator of β_2 is $\beta_2^{\text{TSLS}} = \dfrac{\text{Cov}(Y,\hat{X})}{\text{Var}(\hat{X})} = \dfrac{\text{Cov}(Y,\gamma_1+\gamma_2^{OLS}Z_i)}{\text{Var}(\gamma_1+\gamma_2^{OLS}Z_i)} = \dfrac{\gamma_2^{OLS}\,\text{Cov}(Y,Z_i)}{(\gamma_2^{OLS})^2\,\text{Var}(Z_i)} = \dfrac{\text{Cov}(Y,Z_i)}{\gamma_2^{OLS}\,\text{Var}(Z_i)}$.
Substituting $\gamma_2^{OLS} = \dfrac{Cov(X,Z)}{\text{Var}(Z)}$ into the last expression we get $\hat{\beta}_2^{\text{TSLS}} = \dfrac{Cov(Y,Z_i)}{\dfrac{Cov(X,Z)}{\text{Var}(Z)}\text{Var}(Z_i)} = \dfrac{Cov(Y,Z_i)}{Cov(X,Z)} = \hat{\beta}_2^{IV}$, so $\hat{\beta}_2^{\text{TSLS}} = \hat{\beta}_2^{IV}$.**[3 marks]** **[possible 2 bonus marks]**
[Total 8 marks for e)]

[Total 35 marks for Question 3 + possible bonus marks)].

The International College of Economics and Finance
Econometrics – 2013-2014.
Mid-year exam. December 27.
Suggested solution.

Part 2. (Reading time 20 minutes, working time 1 hour 30 minutes). Answer the first (obligatory) problem and any one of the two problems (2-3).

General information to the problems 1, 2 and 3.

All questions of this examination are based on the same data on the performance of ICEF current 3rd year students during 2011-2013. All equations below reflect the real situation. All questions use the same notations of variables. The number of observations in all regressions of full sample is the same – 100 (except if specified otherwise).

DESCRIPTION OF VARIABLES
(the results of all examinations and courses are measured in points 0-100):
O – The result of the October 2013 mock examination in Econometrics;
MCH – multiple choice result;
FR - free response result;
A – attendance – number of attended seminars in Econometrics in the first module (0, 1, 2, …,8);
H – average grade of submitted home assignments in Econometrics;

UoL Examinations Results
M1 - results of UoL Mathematics 1 (basic) (2nd year of study) ;
M2 - results of UoL Mathematics 2 (2nd year of study);
S1 - results of UoL Statistics 1 (basic) (2nd year of study) ;
S2 - results of UoL Statistics 2 (2nd year of study);

Second Year Mathematical Courses Results
LA – results of Linear Algebra course (2nd year of study);
MATH – results of Mathematics course (2nd year of study);
MOS – results of Methods of Optimal Solutions course (2nd year of study);

Dummy Variables
GZ – belonging to the group taught by Oleg Zamkov, (dummy variable) ;
GB - belonging to the group taught by Elena Bernova, (dummy variable) ;
GC - belonging to the group taught by Vladimir Chernyak, (dummy variable) ;
MALE – dummy variable equal 1 for male, 0 for female (dummy variable).

Standard errors are in parentheses in all estimated equations.

(1) (1a). A researcher considers the following regression (refer to the list of variables on the separate page)

$$\hat{O}_i = -12.2 + 0.26H_i + 0.82A_i + 0.57S2_i \quad R^2 = 0.63 \qquad (1)$$
$$(4.97)\quad (0.05)\quad (0.51)\quad (0.08)$$

Give the interpretation of the model coefficients, and test their significance (including intercept). Explore whether the regression as a whole is significant (state a pair of hypotheses and run appropriate test).

Variables H and $S2$ are significant at ane reasonable level, A is not significant (possibly due to multicollinearity), intercept is significant at 5% level. (t-statistics and critical values for them should be provided – it is possible to use normal distribution).
Each additional attended seminar (keeping H and $S2$ constant increases result of October Econometric exam by 0.82 points (interpretation of other factors is also in this line).
Equation as a whole is significant at any level (F-statistic 53.1 and critical value F(1%, 3, 96)=3.98 for it should be provided, null hypothesis should not include restriction on intercept)

(1b). One of the students remembered that when taking 2nd year UoL Statistics examination many students were very nervous so it is natural to suppose that the results of this exam (variable $S2$) reflect the knowledge of Statistics with some measurement error. How this assumption changes the analysis and interpretation of the equation (2)? Trying to test this assumption the researcher regressed $S2$ on various factors as $S1$, $M1$, LA, $MATH$, MOS, and held the residuals of this auxiliary equation in the variable E. Then she added E to the variables of equation (1) and got equation (2):

$$\hat{O}_i = -23.59 + 0.21H_i + 0.87A_i + 0.79S2_i - 0.51E_i \quad R^2 = 0.66 \qquad (2)$$
$$(5.98)\quad (0.053)\quad (0.49)\quad (0.11)\quad (0.16)$$

Comment on the aim and the logic of the performed procedure and also on the obtained results.

Darbin-Wu-Hausman test for the endogeneity due to the measurement errors. The logic of this test is following: If all explanatory variables in regression are exogenous the estimators of its coefficients are BLUE. If some variables in regression equation are endogenous the estimators of the regression coefficients are no more consistent.
Let we suspect that one of the explanatory variables is endogenous. HAUSMAN TEST for endogeneity (in the version proposed by Davidson and MacKinnon) includes two stages.

1) we regress suspect variable on all exogenous variables and instruments and retrieve the residuals (auxiliary regression).

2) then we re-estimate main regression including the residuals from the first (auxiliary) regression as additional regressors.

If the OLS estimates are consistent, then the coefficient on the first stage residuals should not be significantly different from zero. So the significance of the residuals could be considered as a sign of possible endogeneity of suspect explanatory variable. This is just our case as $t_{E_i} = \frac{0.541}{0.183} = 2.95 > 2.57$

Since endogeneity and stochastic nature of the explanatory variable are respectively confirmed, the researcher should consider the possible consequences of these phenomena. The estimates could be biased (likely downwards as all coefficients could be supposed positive).

If the students were nervous also taking econometric examinations (so the variable $\hat{O}_i$ is also stochastic) the consequences will be different (inefficiency of estimates) as $\hat{O}_i$ is the dependent variable.

(1c). The researcher suspects that the knowledge of the basic Mathematics (variable $M1$) could cause hereroscedastisity of the type $\sigma_u = k \cdot M1$ in the equation

$$M2_i = \beta_1 + \beta_2 S1_i + \beta_2 M1_i + u \qquad (3)$$

To test this she sorts all data in a variable $M1$ (in ascending order) and gets for the upper 40 observations (students with high marks in $M1$) the value of residual sum of squares (RSS) equals 1365.3, while for 25 observations (students with lowest marks $M1$) the value of residual sum of squares (RSS) equals 2608.6. Perform Goldfeld-Quandt test for heteroscedasticity.
She also wants to run White test for heteroscedasticity and gets for corresponding auxiliary equation with cross terms the value of test statistic $n \cdot R^2 = 14.8$. Continue the White test. Compare and comment the results of two tests for heteroscedasticity.
What are the consequences of hereroscedastisity and how they can be mitigated

The Goldfeld-Quandt statistic for this case is $F = \dfrac{2608.6/(25-3)}{1365.3/(40-3)} = 3.213$ while critical value $F(22, 37, 1\%) < F(22, 35, 1\%) = 2.62$ so the null hypothesis of no heteroscedasticity is rejected at 1%.
Auxiliary equation for the White test with cross terms is of the type

$$residuals_i^2 = \gamma_1 + \gamma_2 S1_i + \gamma_3 M1_i + \gamma_4 S1_i^2 + \gamma_5 M1_i^2 + \gamma_6 S1_i \cdot M1_i + v_i,$$

so the statistic $n \cdot R^2$ has χ^2-distribution with 5 degrees of freedom. $\chi^2(5, 1\%) = 15.086$ and $\chi^2(5, 5\%) = 11.070$ so the null hypothesis of no heteroscedasticity is rejected only at 5%. So the Goldfeld-Quandt test based on the assumption of the certain type of heteroscedasticity has an advantage here over more general White test.

(1d) The chance of guessing the correct answer to each question of multiple choice test in Econometrics is equal, for a certain student, to p (each question has 5 alternative answers; p is unknown). The student answered correctly only 4 questions from 12 at the October mock exam in Econometrics, and the teacher believes that the student has no knowledge of Econometrics and relies on luck alone.

(i) Find Maximum Likelihood estimator for the probability p for the student under consideration.

(ii) To what extent the teacher's opinion about the student can be considered as reasonable? (suggest and discuss some procedure and decision rule, no formal test and calculations are expected here).

Let an event is hypothesized to occur with probability *p*. In a sample of *n* observations, it occurred *m* times. We demonstrate that the maximum likelihood estimator of *p* is *m*/*n*.
The probability of the event occurring exactly *m* times in a sample of *n* observations is

$$\frac{n!}{m!(n-m)!} p^m (1-p)^{n-m}.$$

The derivative of this with respect to *p* is

$$\frac{n!}{m!(n-m)!}\left(mp^{m-1}(1-p)^{n-m} - (n-m)p^m(1-p)^{n-m-1}\right)$$

This is equal to zero if

$$m(1-p) - (n-m)p = 0,$$

the condition for which is $p = m/n$. We should check that the second derivative is negative and that we have therefore found a maximum. The second derivative is

$$\frac{n!}{m!(n-m)!}\left(m(m-1)p^{m-2}(1-p)^{n-m} - 2m(n-m)p^{m-1}(1-p)^{n-m-1} + (n-m)(n-m-1)p^m(1-p)^{n-m-2}\right)$$

$$= \frac{n!}{m!(n-m)!} p^{m-2}(1-p)^{n-m-2}\left(m(m-1)(1-p)^2 - 2m(n-m)p(1-p) + (n-m)(n-m-1)p^2\right)$$

The sign of this depends on the sign of

$$\left(m(m-1)(1-p)^2 - 2m(n-m)p(1-p) + (n-m)(n-m-1)p^2\right)$$

Evaluated at $p = m/n$, this is

$$\left(m(m-1)\left(\frac{n-m}{n}\right)^2 - 2m(n-m)\frac{m}{n}\frac{n-m}{n} + (n-m)(n-m-1)\frac{m^2}{n^2} \right).$$

$$= \frac{m(n-m)}{n^2}\big((m-1)(n-m) - 2m(n-m) + m(n-m-1)\big) = -\frac{m(n-m)}{n}.$$

This is negative, so we have indeed chosen the value of p that maximizes the probability of the outcome. Alternative solution could be given using Bernoully distribution.

(Optional). The derived estimator is unbiased.
Let $X_i = 1, 0$ - is the number of successes in the trial $i = 1, ..., n$ so $m = \sum X_i$ then $E\frac{m}{n} = \frac{E\sum X_i}{n} = \frac{\sum EX_i}{n} = \frac{\sum(1\cdot p + 0\cdot(1-p))}{n} = \frac{np}{n} = p$. This means that estimator is unbiased for any finite sample.
Considerations relating efficiency of the derived estimator are also welcomed.

For the situation under consideration one can get estimate $\hat{p}_{MLE} = \frac{m}{n} = \frac{4}{12} = \frac{1}{3}$ while the estimate for the pure guessing is $\hat{p} = \frac{1}{5}$. This probabilities differ but not so much.
(Optional) Some additional considerations are possible. For example one can evaluate probability to get 4 or more successes under assumption that true probability is $\hat{p} = \frac{1}{5} = 0.2$:

$$P_{12}(k \ge 4) = 1 - C_{12}^3 0.2^3 0.8^9 - C_{12}^2 0.2^2 0.8^{10} - C_{12}^1 0.2^1 0.8^{11} - C_{12}^0 0.2^0 0.8^{12} \approx 1 - 0.8 = 0.2$$

This probability is not very high so the teacher's assumption looks not very probable.

(1e) The researcher is studying factors influencing the student's Free Response result in Econometrics (October mock exam). She is especially interested in the study of interconnections between the gender of a student and her/his belonging to the group taught by a certain teacher. So she runs the following equation.

$$\hat{FR}_i = \underset{(5.6)}{5.3} + \underset{(0.7)}{1.4}A_i + \underset{(0.1)}{0.3}H_i + \underset{(6.5)}{15.1}GZ_i + \underset{(5.4)}{13.9}GC_i + \underset{(5.3)}{3.7}MALE_i - \underset{(8.6)}{2.4}MALE_i * GZ_i - \underset{(6.8)}{13.8}MALE_i * GC_i \quad R^2 = 0.5$$

Help the researcher understand the meaning of coefficients of this equation indicating significant coefficients. What is the reference category here? Is the last term of this equation significant? What is its meaning?

All coefficients except the ones of variables $MALE$ and $MALE * GZ$ are significant (at least at 5%). For the coefficient of $MALE * GC$ it could be evaluated as $t = \frac{13.8}{6.8} = 2.03$ while $t(cr., (100-8), 5\%) \approx z(cr.5\%) = 1.96$. The reference category here is the female student belonging to the group tought by Elena Bernova. The expected examination result for such a student who do not attend seminars and do not submit home assignment is 5.317. The result for any female student (with the same number of attended seminars and the same average grade for home assignments) is higher by 13.91 if she is in the group taught by Vladimir Chernyak (coefficient of the variable GC). But this effect is completely canceled for the male student (coefficient of $MALE * GC$ equal to -13.74 so the total effect of combination of two factors GC and $MALE$ is negligable $13.91 - 13.74 = 0.17$, this means that results of male students taught by Elena Bernova and Vladimir Chernyak are approximately the same).

2. A researcher is studying the factors affecting the results of October 2013 mock exam in Econometrics. First she runs two regressions

$$\hat{O}_i = 15.7 + 0.35H_i + 1.27A_i \quad R^2 = 0.45 \qquad (1)$$
$$(3.33)\ (0.063) \quad (0.62)$$

and

$$\hat{O}_i = -12.2 + 0.26H_i + 0.82A_i + 0.57S2_i \quad R^2 = 0.624 \qquad (2)$$
$$(4.97)\ (0.05) \quad (0.51) \quad (0.08)$$

with the covariance matrix (Table 1).

	H	A	$S2$
H	744.4	46.6	148.9
A	46.6	7.79	13.2
$S2$	148.9	13.2	212.2

(2a) Compare the specifications of two equations using various econometric techniques

All coefficients except the one of variable A in equation (2) are significant (at least at 5%). The equation differ only by variable $S2$. Although R-squared for both equations are not very high the both equations are significant (for (1) $F = \frac{0.45/2}{(1-0.45)/97} = 39.7$ with critical value $F(2,97,1\%) = 4.85$, for equation (2) $F = \frac{0.62/3}{(1-0.62)/96} = 52.2$ with critical value $F(3,96,1\%) = 4.01$). The inclusion of $S2$ was reasonable as its t-statistic $t = \frac{0.57}{0.08} = 7.12$ is significant at 1%, moreover as $t = 7.12 > 1$ we have $R^2_{adj}(eq.2) > R^2_{adj}(eq.1)$.

If the true specification corresponds to the equation (2), the omission of the variable $S2$ can lead to omitted variable bias.

(Optional) If the true specification corresponds to the equation (1), what is unlikely in this situatuion, the inclusion of the variable $S2$ can lead to less efficiency of the estimates. One can get correlation matrix

	H	A	$S2$
H	1.00	0.61	0.37
A	0.61	1.00	0.32
$S2$	0.38	0.32	1.00

($Cor(A,S2) = \frac{Cov(A,S2)}{\sqrt{Var(A)\cdot Var(S2)}} = \frac{13.2}{\sqrt{7.79\cdot 212.2}} = 0.32$ and so on)

The correlation coefficients are not very high but probably they are comparable with the correlation coefficients of the dependent variable O and the explanatory variables so the multicollinearity could emerge.

In fact both specification are not correct and more explanatory variables should be used.

(2b) If the researcher believes that correct specification corresponds to equation (2), derive the expression for the omitted variable bias of the estimate of coefficient by the variable *H* under assumption that the variable *S*2 is omitted from the equation.

Please note that for the regression $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + u$ the OLS estimator of β_2 is

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3, Y)\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)\mathrm{Var}(X_3) - [\mathrm{Cov}(X_2, X_3)]^2}$$

Let $Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u$

Let we omit X_4 and run the regression

$$\hat{Y} = b_1 + b_2 X_2 + b_3 X_3$$

$$b_2 = \frac{\mathrm{Cov}(X_2, Y)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3, Y)\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)\mathrm{Var}(X_3) - [\mathrm{Cov}(X_2, X_3)]^2} =$$

$$b_2 = \frac{1}{\Delta}\{\mathrm{Cov}(X_2, \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u)\mathrm{Var}(X_3) - $$
$$- \mathrm{Cov}(X_3, \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + u)\mathrm{Cov}(X_2, X_3)\} =$$

$$b_2 = \frac{1}{\Delta}\beta_2[\mathrm{Var}(X_2)\mathrm{Var}(X_3) - [\mathrm{Cov}(X_2, X_3)]^2 +$$
$$+\frac{1}{\Delta}\beta_3[\mathrm{Cov}(X_2, X_3)\mathrm{Var}(X_3) - \mathrm{Var}(X_3)\mathrm{Cov}(X_2, X_3)] +$$
$$+\frac{1}{\Delta}\beta_4[\mathrm{Cov}(X_2, X_4)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3, X_4)\mathrm{Cov}(X_2, X_3) +$$
$$+\frac{1}{\Delta}[\mathrm{Cov}(X_2, u)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3, u)\mathrm{Cov}(X_2, X_3)]$$

$$b_2 = \frac{1}{\Delta}\beta_2\Delta + \frac{1}{\Delta}\beta_3 \cdot 0 + \frac{1}{\Delta}\beta_4[\mathrm{Cov}(X_2, X_4)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3, X_4)\mathrm{Cov}(X_2, X_3)] +$$
$$+\frac{1}{\Delta}[\mathrm{Cov}(X_2, u)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3, u)\mathrm{Cov}(X_2, X_3)] =$$
$$= \beta_2 + \frac{1}{\Delta}\beta_4[\mathrm{Cov}(X_2, X_4)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3, X_4)\mathrm{Cov}(X_2, X_3)] +$$
$$+\frac{1}{\Delta}[\mathrm{Cov}(X_2, u)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3, u)\mathrm{Cov}(X_2, X_3)]$$

Now take expectation

$$Eb_2 = \beta_2 + \frac{1}{\Delta}\beta_4[\mathrm{Cov}(X_2, X_4)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3, X_4)\mathrm{Cov}(X_2, X_3)] +$$
$$+\frac{1}{\Delta}[\mathrm{Cov}(X_2, Eu)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3, Eu)\mathrm{Cov}(X_2, X_3)] =$$
$$= \beta_2 + \frac{1}{\Delta}\beta_4[\mathrm{Cov}(X_2, X_4)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3, X_4)\mathrm{Cov}(X_2, X_3)]$$

So bias is $\beta_4 \dfrac{[\mathrm{Cov}(X_2, X_4)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3, X_4)\mathrm{Cov}(X_2, X_3)]}{\mathrm{Var}(X_2)\mathrm{Var}(X_3) - [\mathrm{Cov}(X_2, X_3)]^2}$

(2c) Show that the expression $\mathrm{Var}(U)\mathrm{Var}(V)-[\mathrm{Cov}(U,V)]^2$ is always nonnegative for any random variables U and V. Supposing equation (2) is correctly specified evaluate the bias using the data from the Table 1.

It follows from the properties of correlation $-1 \le \mathrm{Cor}(U,V) \le 1$ or $[\mathrm{Cor}(U,V)]^2 \le 1$ or $\frac{[\mathrm{Cor}(U,V)]^2}{\mathrm{Var}(U)\mathrm{Var}(V)} \le 1$ or $\mathrm{Var}(U)\mathrm{Var}(V) \ge [\mathrm{Cov}(U,V)]^2$

Alternatively it could be derived from Cauchy-Schwarz inequality: let $u = U - \bar{U}$, $v = V - \bar{V}$ ($\sum u^2 \sum v^2 \ge (\sum uv)^2$ or $\sum (U-\bar{U})^2 \sum (V-\bar{V})^2 \ge [\sum (U-\bar{U})(V-\bar{V})]^2$, or $\frac{\sum (U-\bar{U})^2}{n-1} \cdot \frac{\sum (V-\bar{V})^2}{n-1} \ge \frac{1}{(n-1)^2}[\sum (U-\bar{U})(V-\bar{V})]^2$ or $\mathrm{Var}(U)\mathrm{Var}(V) \ge [\mathrm{Cov}(U,V)]^2$).

The denominator in the expression for the bias

$$bias(b_2) = \beta_4 \frac{[\mathrm{Cov}(X_2,X_4)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3,X_4)\mathrm{Cov}(X_2,X_3)]}{\mathrm{Var}(X_2)\mathrm{Var}(X_3) - [\mathrm{Cov}(X_2,X_3)]^2}$$

is always positive so the sign of the bias depends on the sign of β_4 and the sign of the numerator $\mathrm{Cov}(X_2,X_4)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3,X_4)\mathrm{Cov}(X_2,X_3)$.

Evaluate bias using data from the equation (2) and the table 1.

$$\hat{O}_i = -12.2 + 0.26H_i + 0.82A_i + 0.57S2_i \quad R^2 = 0.624 \qquad (2)$$
$$\qquad (4.97) \quad (0.05) \qquad (0.51) \quad (0.08)$$

	H	A	$S2$
H	744.4	46.6	148.9
A	46.6	7.8	13.2
$S2$	148.9	13.2	212.2

Using data from the table 1 one obtain

$$bias(b_2) = 0.57 \cdot \frac{148.9 \cdot 7.8 - 13.2 \cdot 46.6}{744.4 \cdot 7.8 - [46.6]^2} = 0.085$$

So the coefficient should be
=0.26+0.085=0.345
In fact in equation (1) $b_2 = 0.35$
So the omitting of variable $S2$ will make the coefficient by H equal to 0.35

(2d) The researcher is trying to understand whether the belonging to the group of a certain teacher is essential for the success at the examination.

$$\hat{O}_i = \underset{(3.4)}{19.3} + \underset{(0.06)}{0.26}H_i + \underset{(0.61)}{1.42}A_i + \underset{(3.74)}{9.44}GZ_i - \underset{(3.09)}{5.37}GB_i \quad R^2 = 0.51 \qquad (3)$$

How the equation (3) would change if the researcher replaces the variable GB_i by the variable GC_i? (Indicate all possible changes and all values that do not change). Is this replacement of any use (in the sense of giving some new information)?

If the researcher replaces the variable GB_i by the variable GC_i the R-squared and significance of the equation (F-statistic) will not change as well as the coefficient of the variables H_i and A_i and their standard errors and significance. The coefficient of GC_i will be opposite to the one of GB_i (equal to plus 5.37) with the same standard error. With the change of reference category the constant term and the coefficient of GZ_i and their standard errors will change (as coefficient of GZ_i will show now the premium in examination result caused by belonging to the group taught by Oleg Zamkov as compared with group taught by Elena Bernova). The last one is the new information that could not be derived from the equation (3) where coefficient of GZ_i shows the premium in examination result caused by belonging to the group taught by Oleg Zamkov as compared with group taught by Vladimir Chernyak. (Optional) In general the changing of reference category gives new information only if the number of reference categories is more than 2.

(2e). Is the factor 'belonging to the group of a certain teacher' in **(2d)** significant? Run the appropriate test (tests).
During discussion one of the participants suggested to use here Chow test. Please comment on this proposal.
Are there any indications that the equation (3) lacks some significant variables? Please explain.

As the factor 'belonging to the group of a certain teacher' is described by two dummy variables GB_i and GC_i the appropriate test is the F-test comparing equations (1) and (3): $F = \dfrac{(0.51-0.45)/2}{(1-0.51)/95} = 5.82$ with $F(2,95,1\%) = 4.85$ so the factor is significant.

Chow test would be more appropriate here as it takes into account also the possible differences in slope dummies $H \cdot GB_i, A \cdot GB_i$, $H \cdot GZ_i, A \cdot GZ_i$ (not included in equation (3)). But the results of Chow test will be different from the F-test above as the last one uses only two dummies.

The R-squared in (3) is only 0.51 so it could be reasonable to look for some additional explanatory variables. The high significance of the dummy variable GZ_i allows to suppose that this variable is important not only in itself, but can be considered as a proxy for some variables describing the general quality of students included in the group taught by Oleg Zamkov (rating, average grade and so on or at least $S2$ used in equation (2)).

(3). It is well known that the profound knowledge of Mathematics helps better understanding Statistics while good experience in theoretical and applied Statistics could help in understanding complex and abstract mathematical ideas. These considerations lead to the attempt to explore the performance of students of ICEF in these subjects simultaneously. The researcher estimates two simultaneous equations

$$S2 = \alpha_1 + \alpha_2 M2 + \alpha_3 S1 + u \qquad (1)$$

$$M2 = \beta_1 + \beta_2 S2 + \beta_3 M1 + v \qquad (2)$$

The results of using OLS for estimation of these equations are

$$\hat{S2}_i = -17.24 + 0.70 M2_i + 0.45 S1_i \quad R^2 = 0.62 \qquad (1^*)$$
$$(6.39) \quad (0.09) \qquad (0.10)$$

$$\hat{M2}_i = 15.03 + 0.44 S2_i + 0.37 M1_i \quad R^2 = 0.60 \qquad (2^*)$$
$$(6.73) \quad (0.07) \qquad (0.10)$$

where $M2$ and $S2$ are supposed to be endogenous while $M1$ and $S1$ (knowledge of basic Mathematics and Statistics) are supposed to be exogenous.

(3a). Why there might be some problems with the using OLS for estimation of both equations? Derive reduced form equations for the system (1)-(2). What can one learn from here on the identification of both equations?

As the dependent variable $S2$ of the equation (1) is included in (2) as an explanatory variable and vice versa the dependent variable $M2$ of the equation (2) is included in (1) as an explanatory variable, the circularity emerges that leads to the fact that the disturbance term u correlates with the explanatory variable $M2$ in equation (1) and the disturbance term v correlates with the explanatory variable $S2$ in equation (2).
Detailed explanation: let '~' means 'correlates with', then from equation (1) $S2 \sim u$ then from equation (2) $S2 \sim M2$ and so $M2 \sim u$, that means that Gauss-Markov conditions are violated for equation (1). The same could be shown for equation (2).
As a consequence of this is the fact that estimates of coefficients of both equations are no most efficient. They are also inconsistent and so biased for the large samples but the proof of this cannot be derived from here and needs more advanced technique.

The system of equations

$$S2 = \alpha_1 + \alpha_2 M2 + \alpha_3 S1 + u \qquad (1)$$

$$M2 = \beta_1 + \beta_2 S2 + \beta_3 M1 + v \qquad (2)$$

is called to be in structural form.
Reduced form equations can be derived from here

$$S2 = \alpha_1 + \alpha_2(\beta_1 + \beta_2 S2 + \beta_3 M1 + v) + \alpha_3 S1 + u$$

$$(1-\alpha_2\beta_2)S2 = \alpha_1 + \alpha_2\beta_1 + \alpha_2\beta_3 M1 + \alpha_3 S1 + u + \alpha_2 v$$

$$S2 = \frac{1}{(1-\alpha_2\beta_2)}[\alpha_1 + \alpha_2\beta_1 + \alpha_2\beta_3 M1 + \alpha_3 S1 + u + \alpha_2 v] \qquad (3)$$

$$M2 = \beta_1 + \beta_2 S2 + \beta_3 M1 + v$$

$$M2 = \beta_1 + \beta_2(\alpha_1 + \alpha_2 M2 + \alpha_3 S1 + u) + \beta_3 M1 + v$$

$$M2 = \frac{1}{(1-\alpha_2\beta_2)}[\beta_1 + \alpha_1\beta_2 + \alpha_3\beta_2 S1 + \beta_3 M1 + v + \beta_2 u] \qquad (4)$$

Both equations (1-2) are exactly identified. In fact equations (3)-(4) contain no circularity, so OLS can be used for estimation of their coefficients, the results of this estimation can be used for obtaining consistent ILS estimates of the coefficients of original structural equations (the exact formulas for these are not expected here).

(3b). Show that OLS estimator of coefficient α_2 of variable $M2$ in equation (1) is inconsistent. Find the expression for the bias on the large sample.

Please note that for the regression $Y = \alpha_1 + \alpha_2 X_2 + \alpha_3 X_3 + u$ the OLS estimator of α_2 is

$$a_2 = \frac{\mathrm{Cov}(X_2, Y)\mathrm{Var}(X_3) - \mathrm{Cov}(X_3, Y)\mathrm{Cov}(X_2, X_3)}{\mathrm{Var}(X_2)\mathrm{Var}(X_3) - [\mathrm{Cov}(X_2, X_3)]^2}$$

If the OLS is used to estimate the first equation

$$\hat{S}2 = a_1 + a_2 M2 + a_3 S1$$

according to general formula

$$a_2 = \frac{\mathrm{Cov}(M2, S2)\mathrm{Var}(S1) - \mathrm{Cov}(S1, S2)\mathrm{Cov}(M2, S1)}{\mathrm{Var}(M2)\mathrm{Var}(S1) - [\mathrm{Cov}(M2, S1)]^2}$$

Decomposition formula for the fixed and random components is

$$a_2 = \alpha_2 + \frac{\mathrm{Cov}(M2, u)\mathrm{Var}(S1) - \mathrm{Cov}(S1, u)\mathrm{Cov}(M2, S1)}{\mathrm{Var}(M2)\mathrm{Var}(S1) - [\mathrm{Cov}(M2, S1)]^2} \qquad (5)$$

Let us evaluate the numerator substituting from reduced form equation (4)

$$\mathrm{Cov}(M2, u)\mathrm{Var}(S1) - \mathrm{Cov}(S1, u)\mathrm{Cov}(M2, S1) =$$

$$= \frac{1}{(1-\alpha_2\beta_2)}[\mathrm{Cov}(\beta_1 + \alpha_1\beta_2 + \alpha_3\beta_2 S1 + \beta_3 M1 + v + \beta_2 u, u)\mathrm{Var}(S1) -$$

$$- \mathrm{Cov}(S1, u)\mathrm{Cov}(\beta_1 + \alpha_1\beta_2 + \alpha_3\beta_2 S1 + \beta_3 M1 + v + \beta_2 u, S1)] =$$

$$= \frac{1}{(1-\alpha_2\beta_2)}\{[\alpha_3\beta_2\mathrm{Cov}(S1, u) + \beta_3\mathrm{Cov}(M1, u) + \mathrm{Cov}(v, u) + \beta_2\mathrm{Var}(u)]\mathrm{Var}(S1) -$$

$$- \mathrm{Cov}(S1, u)[\alpha_3\beta_2\mathrm{Var}(S1) + \beta_3\mathrm{Cov}(M1, S1) + \mathrm{Cov}(v, S1) + \beta_2\mathrm{Cov}(u, S1)]\}$$

Taking the plim of the numerator

$$\mathrm{plim}\{\mathrm{Cov}(M2, u)\mathrm{Var}(S1) - \mathrm{Cov}(S1, u)\mathrm{Cov}(M2, S1)\} =$$

$$= \frac{1}{(1-\alpha_2\beta_2)}\{[\alpha_3\beta_2\sigma_{S1,u} + \beta_3\sigma_{M1,u} + \sigma_{v,u} + \beta_2\sigma_u^2]\sigma_{S1}^2) - \sigma_{S1,u}[\alpha_3\beta_2\sigma_{S1}^2 + \beta_3\sigma_{M1,S1} + \sigma_{v,S1} + \beta_2\sigma_{S1,u}]\} =$$

$$= \frac{\beta_2\sigma_u^2\sigma_{S1}^2}{(1-\alpha_2\beta_2)}$$

as according the assumption of the model $\sigma_{S1,u} = 0,\ \sigma_{v,S1} = 0,\ \sigma_{M1,u} = 0,\ \sigma_{v,u} = 0$

Let us evaluate the denominator in (5) substituting from reduced form equation (4)

$$\mathrm{Var}(M2)\mathrm{Var}(S1) - [\mathrm{Cov}(M2, S1)]^2 =$$

$$= \frac{1}{(1-\alpha_2\beta_2)^2}\{\mathrm{Var}(\beta_1 + \alpha_1\beta_2 + \alpha_3\beta_2 S1 + \beta_3 M1 + v + \beta_2 u)\mathrm{Var}(S1) -$$

$$- [\mathrm{Cov}(\beta_1 + \alpha_1\beta_2 + \alpha_3\beta_2 S1 + \beta_3 M1 + v + \beta_2 u, S1)]^2\} =$$

Taking the plim of the numerator and omitting zero terms

$$\mathrm{plim}\{\mathrm{Var}(M2)\mathrm{Var}(S1) - [\mathrm{Cov}(M2, S1)]^2\} =$$

$$= \frac{1}{(1-\alpha_2\beta_2)^2}\{[\alpha_3^2\beta_2^2\sigma_{S1}^2 + \beta_3^2\sigma_{M1}^2 + \sigma_v^2 + \beta_2^2\sigma_u^2 + 2\sigma_{S1,M1}]\sigma_{S1}^2 - [\alpha_3\beta_2\sigma_{S1}^2 + \beta_3\sigma_{S1,M1}]^2\}$$

Finally $$p\lim a_2 = \alpha_2 + \frac{(1-\alpha_2\beta_2)\beta_2\sigma_u^2\sigma_{S1}^2}{[\alpha_3^2\beta_2^2\sigma_{S1}^2 + \beta_3^2\sigma_{M1}^2 + \sigma_v^2 + \beta_2^2\sigma_u^2 + 2\sigma_{S1,M1}]\sigma_{S1}^2 - [\alpha_3\beta_2\sigma_{S1}^2 + \beta_3\sigma_{S1,M1}]^2} \qquad (5)$$

(3c) Show that the expression $\mathrm{Var}(U)\mathrm{Var}(V)-[\mathrm{Cov}(U,V)]^2$ is always nonnegative for any random variables U and V. Based on this evaluate the sign of the bias and determine on what factors it depends.

It follows from the properties of correlation $-1 \le \mathrm{Cor}(U,V) \le 1$ or $[\mathrm{Cor}(U,V)]^2 \le 1$ or $\frac{[\mathrm{Cor}(U,V)]^2}{\mathrm{Var}(U)\mathrm{Var}(V)} \le 1$ or $\mathrm{Var}(U)\mathrm{Var}(V) \ge [\mathrm{Cov}(U,V)]^2$

Alternatively it could be derived from Cauchy-Schwarz inequality: let $u = U - \bar{U}$, $v = V - \bar{V}$ ($\sum u^2 \sum v^2 \ge (\sum uv)^2$ or $\sum (U-\bar{U})^2 \sum (V-\bar{V})^2 \ge [\sum (U-\bar{U})(V-\bar{V})]^2$, or $\frac{\sum (U-\bar{U})^2}{n-1} \cdot \frac{\sum (V-\bar{V})^2}{n-1} \ge \frac{1}{(n-1)^2}[\sum (U-\bar{U})(V-\bar{V})]^2$ or $\mathrm{Var}(U)\mathrm{Var}(V) \ge [\mathrm{Cov}(U,V)]^2$).

The denominator in the expression

$$p\lim a_2 = \alpha_2 + \frac{(1-\alpha_2\beta_2)\beta_2\sigma_u^2\sigma_{S1}^2}{[\alpha_3^2\beta_2^2\sigma_{S1}^2 + \beta_3^2\sigma_{M1}^2 + \sigma_v^2 + \beta_2^2\sigma_u^2 + 2\sigma_{S1,M1}]\sigma_{S1}^2 - [\alpha_3\beta_2\sigma_{S1}^2 + \beta_3\sigma_{S1,M1}]^2}$$

is always positive (as the limiting value of the positive expression $\mathrm{Var}(M2)\mathrm{Var}(S1)-[\mathrm{Cov}(M2,S1)]^2$) so the sign of the bias depends on the sign of β_4 and the sign of the numerator.

The factor $(1-\alpha_2\beta_2)$ is generally positive, so the sign of the bias depends on the sign of β_2. In our case β_2 is presumably positive from the meaning of variables so the bias is positive.

(3d) At the seminar one of the participants indicated that the knowledge of basic mathematics (measured by $M1$) is also essential in the understanding of second year Statistics. How this proposal changes the situation with the identification of both equations? Another participant suggested that for the sake of symmetry it would be logical to include the variable $S1$ into the second equation. Comment on what might be the consequences of these proposals for the identification of both equations.

Both equations of the original system of structural equations

$$S2 = \alpha_1 + \alpha_2 M2 + \alpha_3 S1 + u \qquad (1)$$

$$M2 = \beta_1 + \beta_2 S2 + \beta_3 M1 + v \qquad (2)$$

are exactly identified as exogenous variable $M1$ can be used as the instrument for $M2$ in the first equation, and exogenous variable $S1$ can be used as the instrument for $S2$ in the second equation.

The first proposal leads to the system of equations

$$S2 = \alpha_1 + \alpha_2 M2 + \alpha_3 S1 + \alpha_4 M1 + u \qquad (1\text{-}1)$$

$$M2 = \beta_1 + \beta_2 S2 + \beta_3 M1 + v \qquad (2\text{-}1)$$

Using instrumental variable approach or the order condition we can see that first equation becomes underidentified as variable $M1$ cannot be used any more as the instrument for $M2$, while second equation remains exactly identified.

The second proposal leads to a complete collapse

$$S2 = \alpha_1 + \alpha_2 M2 + \alpha_3 S1 + \alpha_4 M1 + u \qquad (1\text{-}2)$$

$$M2 = \beta_1 + \beta_2 S2 + \beta_3 M1 + \beta_4 S1 + v \qquad (2\text{-}2)$$

as both equations become underidentidied, so it is impossible to obtain consistent estimates of their parameters.

(3e) Suppose that both suggestions were rejected. The researcher instead uses the result of three mathematical courses of the second year (LA, $MATH$, MOS , - refer to the list of variables), for estimation regression of $M2$ on $M1$, LA, $MATH$, and MOS, then saves the estimated values of $M2$ in variable $M2F$, and then runs the following regression:

$$\hat{S2}_i = -30.1 + 0.99 M2F_i + 0.29 S1_i \qquad R^2 = 0.62 \qquad (3)$$
$$(7.36) \quad (0.12) \qquad (0.11)$$

Comment on the aim and the logic of the performed procedure and also on the results obtained.

This is a typical example of using different instruments outside of the system to get consistent estimates of parameters of the equation using Two Stage Least Squares (TSLS) method. This suggestion leads to the system of equations

$$S2 = \alpha_1 + \alpha_2 M2 + \alpha_3 S1 + u \qquad (1\text{-}3)$$
$$M2 = \beta_1 + \beta_2 S2 + \beta_3 M1 + \beta_4 LA + \beta_5 MATH + \beta_6 MOS + v \qquad (2\text{-}3)$$

Second equation remains exactly identified while first equation is overidentified as four different instruments compete to get consistent estimates. Running regression (3) and memorizing the estimated values of dependent variable the researcher performs the first step of TSLS (creating the most efficient instrument $M2F$ combining different possible instruments with optimal weights). Then this instrument is used to get consistent estimates of parameters for the first equation (second step of TSLS).

Of course, the efficiency of this ‘most efficient’ instrument can be nevertheless usually not very high. It can be illustrated by comparison of the standard errors of the equation (1*) and (3)

$$\hat{S2}_i = -17.24 + 0.70 M2_i + 0.45 S1_i \quad R^2 = 0.62 \qquad (1^*)$$
$$(6.39) \quad (0.09) \qquad (0.10)$$

$$\hat{S2}_i = -30.1 + 0.99 M2F_i + 0.29 S1_i \qquad R^2 = 0.62 \qquad (3)$$
$$(7.36) \quad (0.12) \qquad (0.11)$$

One can see that estimates of the coefficients have changed and standard errors became greater.

The International College of Economics and Finance
Econometrics – 2014-2015.
Mid-year exam. December 26.
Marking scheme.

Part 2. (Reading time 20 minutes, working time 1 hour 30 minutes). Candidates should answer SIX of the following SEVEN questions: ALL of the questions in Section A (10 marks each, 50 marks total) and ONE question from Section B (25 marks each). Candidates are strongly advised to divide their time accordingly.

SECTION A

Answer **ALL** questions from this section.

1. The researcher regresses the natural log of expenditure on beer at 2010 prices ($beer_t$) on the natural log of total household expenditure at 2010 prices (exp_t), the natural log of the price of beer relative to all consumer prices (pb_t,) and the natural log of the price of alcoholic drinks excluding beer relative to all consumer prices (pa_t), gave the following results:

$$\underset{(1.387)}{beer_t = -5.272} + \underset{(0.114)}{1.266exp_t} - \underset{(0.446)}{0.989\,pb_t} - \underset{(0.144)}{0.412\,pa_t} + e_t$$

where e_t is the estimated residual, standard errors are in brackets, the sample size is 14, and $R^2 = 0.906$. All assumptions of the model B are assumed to be satisfied.

(a) Use confidence interval method to test whether coefficients of pb_t and pa_t are significant. Explain the meaning of confidence interval and the logic of your decision.

Solution: The 95% confidence interval for β_{pb_t} under ($14-4=10$ degrees of freedom) is $-0.989 \pm 2.228 \times 0.496$ - $[-1.982688; 0.004688]$. The confidence interval gives an interval in which we are 95% sure of containing the true value of the parameter. It constitutes of all null hypotheses compatible with the data. As it includes zero hypothesis $H_0 : \beta_{pb_t} = 0$ is not rejected. For β_{pa_t} we get $-0.412 \pm 2.228 \times 0.144$ - $[-0.821952; -0.091168]$ (but $-0.412 + 3.169 \times 0.144 = 0.044$) - significant only at 5%.

(b) Do your conclusions changes if it can be assumed that coefficients of pb_t and pa_t cannot be positive? Can one of the variables pb_t or pa_t (or both) be dropped from the equation on the basis of your decision?

If coefficients of pb_t and pa_t cannot be positive we can use one-sided values 1.812 instead of 2.228 (5%) and 2.764 instead of 3.169. So for $H_0 : \beta_{pb_t} = 0$, $H_a : \beta_{pb_t} < 0$ we get $t = \frac{-0.989}{0.446} = 2.217 > 1.812$ (significant), and for $H_0 : \beta_{pa_t} = 0$, $H_a : \beta_{pa_t} < 0$ we get $t = \frac{-0.412}{0.144} = 2.8611 > 2.764$ - now significant at 1%.

If the specification of the equation under consideration is correct (what is probably true as all its coefficients are significant) the dropping of any variable could lead to the omitted variable bias.

(c) The researcher suspects that the theoretical coefficients of pb_t and pa_t are equal in value. How to test this hypothesis against alternative hypothesis that they are not equal? What information is neaded additionally for this? Is it possible to test this hypothesis against the alternative that the coefficient of pb_t is greater than the coefficient of pa_t? What is economic meaning of these hypotheses?

First let us wright theoretical equation, corresponding estimated equation above

$$beer_t = \beta_1 + \beta_2 exp_t + \beta_3 pb_t + \beta_4 pa_t + u_t$$

The restriction is $\beta_3 = \beta_4$ with the meaning that price elasticities of beer consumption and consumption of other beverages are equal.

To test this restriction the researcher needs to run equation

$$beer_t = \beta_1 + \beta_2 exp_t + \beta_3 (pb_t + pa_t) + u_t$$

memorize the R_R^2 of restricted equation and compare it with $R_U^2 = 0.906$ of initial unrestricted equation using F-test with test statistic $F = \dfrac{R_U^2 - R_R^2}{(1-R_U^2)/(14-4)}$. If $F > F_{crit}$ the null hypothesis $H_0 : \beta_3 = \beta_4$ is rejected in favor of alternative $H_a : \beta_3 \neq \beta_4$.

Under one sided alternative $H_a : \beta_3 > \beta_4$ the F-test is useless. Here we should use reparametrization of the original equation

$$beer_t = \beta_1 + \beta_2 exp_t + \beta_3 pb_t + \beta_4 pa_t + u_t$$

as

$$beer_t = \beta_1 + \beta_2 exp_t + (\beta_3 - \beta_4) pb_t + \beta_4 (pb_t + pa_t) + u_t$$

or

$$beer_t = \beta_1 + \beta_2 exp_t + \beta \cdot pb_t + \beta_4 (pb_t + pa_t) + u_t \quad \text{where } \beta = \beta_3 - \beta_4$$

After running last equation we use une sided t-test for coefficient β: $H_0 : \beta = 0$ versus $H_0 : \beta > 0$.

2. Let X be a random variable distributed with mean 0 and variance σ^2. Let $X_1, X_2, ..., X_T$ be the identically and independently distributed random sample from the distribution of X.

(a) Show that $S^2 = \dfrac{\sum_{t=1}^{T}(X_t - \overline{X})^2}{T}$; where $\overline{X} = \dfrac{\sum_{t=1}^{T} X_t}{T}$ is a biased estimator of, $\sigma^2 < \infty$. How your proof would change, if you omit the assumption $EX = 0$?

Solution: You are required to take expectation of S^2 and show that it is not equal to σ^2. The answer is:

$$\mathrm{E}(S^2) = \frac{\mathrm{E}\sum (X_t - \overline{X})^2}{T} = \frac{\sum \mathrm{E}(X_t - \overline{X})^2}{T}$$

We can write:

$$\sum \mathrm{E}(X_t - \overline{X})^2 = \sum \mathrm{E}(X_t^2) - T\,\mathrm{E}(\overline{X}^2)$$

$$\mathrm{E}(X_t^2) = \mathrm{var}(X_t) + [\mathrm{E}(X)]^2 = \sigma^2;\ \mathrm{E}(X) = 0$$

$$\mathrm{E}(\overline{X}^2) = \mathrm{var}(\overline{X}) + [\mathrm{E}(\overline{X})]^2 = \frac{\sigma^2}{T};\ \mathrm{E}(\overline{X}) = 0.$$

Hence:

$$\sum \mathrm{E}(X_t - \overline{X})^2 = \sum \sigma^2 - T\left[\frac{\sigma^2}{T}\right] = T\sigma^2 - \sigma^2 = (T-1)\sigma^2 .$$

Therefore:

$$E(S^2) = \frac{(T-1)\sigma^2}{T} \neq \sigma^2 \quad \Longrightarrow \quad S^2 \text{ is a biased estimator of } \sigma^2 .$$

If we omit the assumption that $EX = 0$ almost nothing will change, the variance does not change with the changing of the population mean. To keep use the same proof, one need to use centered values $x_t = X_t - \overline{X}$ (instead of X_t) with zero mean and the same variance σ^2.

(b) As is known, in the case of normally distributed random variable X maximum likelihood estimator of σ^2 is $S_T^2 = \frac{\sum_{t=1}^{T}(X_t - \overline{X})^2}{T}$. What advantages have this estimator as compared to the unbiased estimator of the variance $s_{T-1}^2 = \frac{\sum_{t=1}^{T}(X_t - \overline{X})^2}{T-1}$ (no proof expected) ?

The estimator $S^2 = \frac{\sum_{t=1}^{T}(X_t - \overline{X})^2}{T}$ is (weakly) consistent ($\operatorname*{plim}_{T\to\infty} S^2 = \frac{\sum_{t=1}^{T}(X_t - \overline{X})^2}{T} = \sigma^2$) and asymptotically unbiased ($\operatorname*{plim}_{T\to\infty} \frac{S_T^2}{s_{T-1}^2} = 1$). . Moreover it is asymptotically efficient (having a smaller mean square error than the unbiased estimator in large samples).

3. The researcher is interested in estimation of the following models:

$$y_t = \alpha + \beta x_t + z_t + u_t \qquad (1)$$

$$y_t = \alpha + \beta x_t - \beta z_t + v_t \qquad (2)$$

(a) Find the least squares estimates of β, indicating necessary transformation of the data.

HINT: Note that these models can actually be rewritten as simple regression models with only one explanatory variable with different restrictions

Both regressions are in fact simple linear regression models, so in both cases one should use conventional estimator of the type $\hat{\beta} = \dfrac{\text{Cov}(Y,X)}{\text{Var}(X)}$. But the transformations of the data needed to reduce problem to simple regression model, are different, hence estimators will be different.

For specification (1) one can rewrite the model in the form $y_t - z_t = \alpha + \beta x_t + u_t$ and then estimate the model $y_t^* = \alpha + \beta x_t + u_t$, where $y_t^* = y_t - z_t$.

For specification (2) we estimate the model $y_t = \alpha + \beta x_t^* + u_t$ where $x_t^* = x_t - z_t$.

For (a) we get $\hat{\beta} = \dfrac{\text{Cov}(Y-Z,X)}{\text{Var}(X)} = \dfrac{\text{Cov}(Y,X) - \text{Cov}(Z,X)}{\text{Var}(X)}$

For (b) we get $\hat{\beta} = \dfrac{\text{Cov}(Y,X-Z)}{\text{Var}(X-Z)} = \dfrac{\text{Cov}(Y,X) - \text{Cov}(Z,X)}{\text{Var}(X) + \text{Var}(Z) - 2\,\text{Cov}(Z,X)}$.

(b) Evaluate the mean and the variance of these estimators given that x_t and z_t are deterministic sequences and $u_t \sim iid(0,\sigma^2)$, $v_t \sim iid(0,\sigma^2)$. How to choose between (1) and (2)? Under what conditions the estimator of β from (1) is preferable to that from (2)?

Both regressions are the restricted versions of the general model $y_t = \alpha + \beta x_t + \gamma z_t + u_t$

The restriction for the model (1) is $\gamma = 1$; for (2) it is $\gamma = -\beta$. One first should test each restriction and estimate the general model if both restrictions are invalid, the model with valid restriction if only one is valid (which is preferable for this case).

In a valid model (1) or (2) the estimator of β is unbiased, and it is biased if the restriction leading to the model is invalid.

If both restrictions are valid, any of them can be estimated, but the one with smaller variance is preferable**.** The variance for the estimator of regression coefficient is given by expression $\sigma_{\hat{\beta}}^2 = \dfrac{\sigma_u^2}{n\,\text{Var}(X)}$. For the model (1) (if it is valid) $\sigma_{\hat{\beta}}^2 = \dfrac{\sigma_u^2}{n\,\text{Var}(x_t)}$; for (2) $\sigma_{\hat{\beta}}^2 = \dfrac{\sigma_u^2}{n\,\text{Var}(x_t - z_t)}$. So if both models are valid, the first one is preferable if $\text{Var}(x_t) > \text{Var}(x_t - z_t)$, and the second one otherwise.

To choose between models it is also possible to use $MSE = (bias(\hat{\beta}))^2 + \sigma_{\hat{\beta}}^2$.

4. Suppose the true relationship between Y and X looks like $Y_i = \beta X_i + u_i$. In other words Y_i is distributed around $\beta X_i + u_i$ according to the values of u_i. Let's assume that $u_i = Y_i - \beta X_i$ is normally distributed with zero mean and variance σ^2, its probability density function is $f(u) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{u}{\sigma}\right)^2}$. Derive and compare OLS and ML estimators of β. Assume that n sample values of the disturbance term are independent and not correlated with X ,

Solution

The OLS estimator could be derived as a consequence of minimizing the sum of squares of errors, i.e. minimizing $RSS = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} [Y_i - \hat{\beta} X_i]^2$. Differentiating with respect to β and equating to zero gives $\frac{\partial RSS}{\partial \hat{\beta}} = 2\sum_{i=1}^{b} (Y_i - \hat{\beta} X_i)(-X_i) = 0$ from which it follows that $\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}$. A good answer might go on to show that the stationary point where the derivative =0 is a minimum: let us take the second derivative $\frac{\partial^2 RSS}{\partial \hat{\beta}^2} = \frac{\partial}{\partial \hat{\beta}}\left(-2\sum_{i=1}^{n} (Y_i X_i - \hat{\beta} X_i^2) \right) = 2\sum_{i=1}^{n} X_i^2 > 0$ what indicated that here we have point of minima.

To derive ML estimator we should first construct the likelihood function

$$L(\beta, \sigma \mid u_1, u_2.,..u_n) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{Y_1 - \beta X_1}{\sigma}\right)^2} \times \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{Y_2 - \beta X_2}{\sigma}\right)^2} \times \ldots \times \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{Y_n - \beta X_n}{\sigma}\right)^2}.$$

Log likelihood function becomes:

$$\log L(\beta, \sigma \mid u_1, u_2.,..u_n) = \sum_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{Y_i - \beta X_i}{\sigma}\right)^2} = n\log\frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{Y_i - \beta X_i}{\sigma}\right)^2.$$

The last expression could be represented through RSS (with 'minus' sign)

$$\log L(\beta, \sigma \mid u_1, u_2.,..u_n) = n\log\frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^3}\sum_{i=1}^{n}(Y_i - \beta X_i)^2 = n\log\frac{1}{\sigma\sqrt{2\pi}} - \frac{1}{2\sigma^3} RSS$$

From the obtained expression, maximization of the log likelihood function implies minimization of RSS for choosing estimator of β, Hence, ML estimator of β coincide with the OLS one.

5. Consider a model

$$Y_i = \beta_1 + \beta_2 X_i + u_i;\ i = 1, 2, ..., n,$$

where Y_i is a binary variable that takes the value of 1 if the event takes place and 0 otherwise, and $\mathrm{E}(u_i) = 0$ for $i = 1, 2, ..., n$.

(a) Explain why the problem of heteroscedasticity arises if the model above is estimated by ordinary least squares (OLS).

Solution:

Linear probability model is a model in which dependent variable is binary which takes value 1 if the event occurs and 0 if it does not. It is estimated by the ordinary least squares (OLS).

Let the model be:

$$Y_i = \beta_0 + \beta_1 X_i + u_i;\ i = 1, 2, ..., n \qquad (1)$$

where:

$$Y_i = \begin{cases} 1 \text{ if event occurs} \\ 0 \text{ if not} \end{cases}$$

As $\mathrm{E}(u_i) = 0$, then:

$$\mathrm{E}[Y_i \mid X_i] = \beta_0 + \beta_1 X_i. \qquad (2)$$

Also:

$$\mathrm{E}[Y_i \mid X_i] = 1 \cdot P(Y_i = 1) + 0 \cdot P(Y_i = 0) = P(Y_i = 1) = P_i \qquad (3)$$

From (ii) and (iii):

$$\mathrm{E}[Y_i \mid X_i] = \beta_0 + \beta_1 X_i = P_i$$

hence, we can interpret $\mathrm{E}[Y_i \mid X_i] = \beta_0 + \beta_1 X_i$ as the probability that the event will occur given X_i.

If we denote $\hat{\beta}_0$ and $\hat{\beta}_1$ as an estimate of β_0 and β_1, then we can write:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i = \hat{P}_i \qquad (4)$$

as the estimated probability that the event will occur.

As Y_i takes only two values 1 or 0, therefore u_i can take only two values $1 - \beta_0 - \beta_1 X_i$ when $Y_i = 1$ and $\beta_0 + \beta_1 X_i$ when $Y_i = 0$. Based on this we can write the probability distribution of u_i as:

Y_i	u_i	$f(u_i)$
1	$1 - \beta_0 - \beta_1 X_i$	$\beta_0 + \beta_1 X_i$
0	$-\beta_0 - \beta_1 X_i$	$1 - \beta_0 - \beta_1 X_i$

This probability distribution also satisfies the assumption that:

$$\mathrm{E}(u_i) = (1 - \beta_0 - \beta_1 X_i)(\beta_0 + \beta_1 X_i) + (-\beta_0 - \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i) = 0$$

We can write $\mathrm{var}(u_i)$ as:

$$\begin{aligned}
&\mathrm{var}(u_i) = \mathrm{E}(u_i^2) = \\
&= (1 - \beta_0 - \beta_1 X_i)^2 (\beta_0 + \beta_1 X_i) + (-\beta_0 - \beta_1 X_i)^2 (1 - \beta_0 - \beta_1 X_i) = \\
&= (1 - \beta_0 - \beta_1 X_i)(\beta_0 + \beta_1 X_i)[(1 - \beta_0 - \beta_1 X_i) + (\beta_0 + \beta_1 X_i)] = \\
&= (\beta_0 + \beta_1 X_i)(1 - \beta_0 - \beta_1 X_i) = \mathrm{E}(Y_i)[1 - \mathrm{E}(Y_i)] = \\
&= P_i(1 - P_i);\ \text{for all } i = 1, 2, ..., n.
\end{aligned}$$

The function $\text{var}(u_i) = P_i(1-P_i)$ can take different values from 0 to 0.25 depending on the value of $P_i \in [0,1]$ of each observation. Hence the disturbance term is heteroscedastic. This will make OLS estimators inefficient.

In many cases the estimated probability $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$ will be negative or greater than 1.

(b) If heteroscedasticity was a major shortcoming of the model how would you estimate the model by weighted least squares where the weights are the estimated standard deviations? Discuss the advantages and disadvantages of this procedure.

Solution:

b) Weighted least squares:

We can see from (4), that estimator of P_i is $\hat{P}_i = \hat{Y}_i$, therefore $\hat{Y}_i(1-\hat{Y}_i)$ can be used as an estimator of:

$$\text{var}(u_i) = \text{E}(Y_i)[1-\text{E}(Y_i)] = P_i(1-P_i)\,.$$

Weights can be obtained as:

$$W_i = [\hat{Y}_i(1-\hat{Y}_i)]^{1/2}\,.$$

Divide (1) by W_i and apply OLS to:

$$\frac{Y_i}{W_i} = \frac{\beta_0}{W_i} + \beta_1\frac{X_i}{W_i} + \frac{u_i}{W_i};\ i = 1, 2, ..., n$$

obtain the *WLS* estimator of β_0 and β_1. This will give an efficient estimator.

(***Optional material***) Some problems, connected with this method could be indicated. In practice the estimated variance of u_i, $\hat{Y}_i(1-\hat{Y}_i)$ may be negative as again the estimated probability $\hat{Y}_i$ may be negative or greater than l. The obvious correction of the estimated negative probability is to constrain estimated probabilities within $[0,1]$ interval. If we do this we might predict an occurrence with probability 1, when it is possible that it might not occur or we might predict an occurrence with probability 0 when it might actually occur. The estimation process may give unbiased estimates but predictions obtained from it will be biased.

SECTION B

Answer **ONE** question from this section (**6 OR 7**).

6. A researcher investigating the shadow economy (illegal economy) using international cross-section data for 40 countries hypothesizes that consumer expenditure on shadow goods and services, *q*, is related to total consumer expenditure, *z*, by the relationship

$$q = \alpha + \beta z + v$$

where *v* is a disturbance term which satisfies the Gauss-Markov conditions. Both variables *q* and *z* are measured with error, and from the meaning of variables of this model *q* is part of *z* , so any error in the estimation of *q* affects the estimate of *z* **by the same amount**. Hence

$$y_i = q_i + w_i$$

and

$$x_i = z_i + w_i$$

where y_i is the estimated value of q_i, x_i is the estimated value of z_i, and w_i is the measurement error **affecting both variables** in observation *i*. It is assumed that the expected value of *w* is zero and that *v* and *w* are distributed independently of *z* and of each other. Note since shadow expenditure is a component of total consumer expenditure, β will lie between 0 and 1.

(a) Derive an expression for the large-sample bias in the estimate of β when Ordinary Least Squares is used to regress y_i on x_i, and determine its sign if this is possible.

Solution and marking

a) The relationship between the observed variables is

$$(y - w) = \alpha + \beta(x - w) + v$$

so

$$y = \alpha + \beta x + v + (1-\beta)w = \alpha + \beta x + u$$

where

$$u = v + (1-\beta)w$$

If *y* is regressed on *x* using OLS,

$$b_{OLS} = \frac{\mathrm{Cov}(x,y)}{\mathrm{Var}(x)} = \frac{\mathrm{Cov}(x,[\alpha+\beta x+u])}{\mathrm{Var}(x)}$$
$$= \frac{\mathrm{Cov}(x,\alpha)+\mathrm{Cov}(x,\beta x)+\mathrm{Cov}(x,u)}{\mathrm{Var}(x)} =$$
$$= \beta + \frac{\mathrm{Cov}(x,u)}{\mathrm{Var}(x)}$$

It is not possible to obtain an expression for the expected value of b_{OLS} because both the numerator and the denominator of the error term are functions of *w*, so we will investigate its plim instead.

$$\mathrm{plim}\,\mathrm{Cov}(x,u) = \mathrm{plim}\,\mathrm{Cov}([z+w],[v+(1-\beta)w])$$
$$= \mathrm{plim}\,\mathrm{Cov}(z,v) + (1-\beta)\,\mathrm{plim}\,\mathrm{Cov}(z,w) +$$
$$+ \mathrm{plim}\,\mathrm{Cov}(w,v) + (1-\beta)\,\mathrm{plim}\,\mathrm{Var}(w)$$

The first three plims are zero because *v* and *w* are distributed independently of *z* and each other, so

$$\mathrm{plim}\,\mathrm{Cov}(x,u) = (1-\beta)\sigma_w^2$$
$$\mathrm{plim}\ \mathrm{Var}(x) = \mathrm{plim}\ \mathrm{Var}(z+w) =$$
$$= \mathrm{plim}\ \mathrm{Var}(z) + \mathrm{plim}\ \mathrm{Var}(w) + 2\,\mathrm{plim}\ \mathrm{Cov}(z,w) =$$

$$= \sigma_z^2 + \sigma_w^2$$

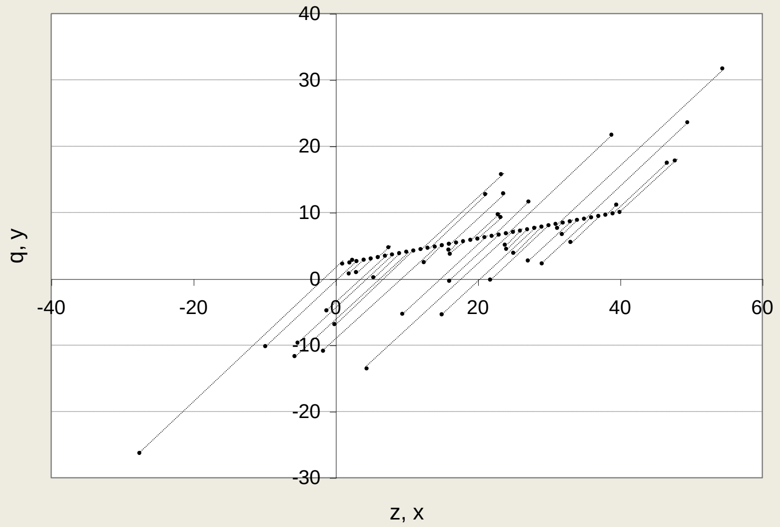
Hence

$$\operatorname{plim} b_{OLS} = \beta + (1-\beta)\frac{\sigma_w^2}{\sigma_z^2 + \sigma_w^2}$$

Since β will lie between 0 and 1, the bias will be positive.

(b) Show that large sample value of the slope of the regression line is the weighted average of the true slope, β, and unity: $\operatorname{plim} b_{OLS} = C_1 \cdot \beta + C_2 \cdot 1$; $C_1 + C_2 = 1$. Find C_1 and C_2. Explain the possible meaning of this relation formally, economically and graphically.

From the result above $\operatorname{plim} b_{OLS} = \beta + (1-\beta)\frac{\sigma_w^2}{\sigma_z^2 + \sigma_w^2}$ we get immediately $\operatorname{plim} b_{OLS} = \frac{\sigma_z^2}{\sigma_z^2 + \sigma_w^2}\beta + \frac{\sigma_w^2}{\sigma_z^2 + \sigma_w^2}$. Hence the slope of the regression line will be a compromise between the true slope, β, and unity. Thus plim b_{OLS} is a weighted average of β and unity, the weights being the variances of z and w. The graph below shows plots the points (q, z) and (y, x) for the first sample, with each (q, z) point linked to the corresponding (y, x) point.



The diagram shows how the measurement error causes the observations to be displaced along 45° lines (slope of these lines is unity). The greater is the variance of the error term w_i the more displacement is observed along these lines with the slope 1, and so the greater if the weight of 1.

(c) The researcher is worried of the fact that consumer expenditure on the shadow economy were systematically underestimated, so the expected value of the measurement error being not zero but rather negative? He is also worried of the fact that the analysis could be affected by positive correlation of *w* with *z*, as countries with large *z* tend to have larger measurement errors *w*. Comment.

The results would not be affected by the expected value of the measurement error term. The analysis of the large-sample bias in fact use no assumption concerning $E(w)$.
The second modification is in fact essential for analysis. We would now have

$$\operatorname{plim}\operatorname{Cov}(x,u) = \operatorname{plim}\operatorname{Cov}([z+w],[v+(1-\beta)w]) =$$
$$= \operatorname{plim}\operatorname{Cov}(z,v) + (1-\beta)\operatorname{plim}\operatorname{Cov}(z,w) + \operatorname{plim}\operatorname{Cov}(w,v) + (1-\beta)\operatorname{plim}\operatorname{Var}(w)$$

Hence

$$\text{plim}\, b_{OLS} = \beta + (1-\beta)\frac{\sigma_w^2 + \sigma_{w,z}}{\sigma_z^2 + \sigma_w^2 + 2\sigma_{w,z}}$$

where $\sigma_{w,z}$ is $\text{plim}\,\text{Cov}(w,z)$. The bias could either increase or decrease, depending on the sign and the value of β.

(d) Trying to overcome consequences of bias caused by measurement errors the researcher decided to use disposable personal income, I as an instrument for total consumer expenditure, z, assuming that I correlates with z but not correlates with v and w. Comment providing necessary proofs, taking into account that consumer expenditures on shadow goods and services, q, still are under measurement errors w.

The idea is quite reasonable, as instrumental variable I allows to get unbiased estimator for β. Taking in mind that both q and z are unobservable we get

$$b_{IV} = \frac{\text{Cov}(I,y)}{\text{Cov}(I,x)} = \frac{\text{Cov}(I,q+w)}{\text{Cov}(I,z+w)} = \frac{\text{Cov}(I,\alpha+\beta z+v+w)}{\text{Cov}(I,z+w)} =$$

$$= \frac{\text{Cov}(I,\alpha)+\text{Cov}(I,\beta z)+\text{Cov}(I,v+w)}{\text{Cov}(I,z+w)} = \frac{0+\beta\cdot\text{Cov}(I,z)+\text{Cov}(I,v)+\text{Cov}(I,w)}{\text{Cov}(I,z)+\text{Cov}(I,w)}$$

Taking probability limit we get

$$\text{plim}\, b_{IV} = \frac{\beta\cdot\text{plim}\,\text{Cov}(I,z)+\text{plim}\,\text{Cov}(I,v)+\text{plim}\,\text{Cov}(I,w)}{\text{plim}\,\text{Cov}(I,z)+\text{plim}\,\text{Cov}(I,w)} =$$

$$= \frac{\beta\cdot\sigma_{I,z}+\sigma_{I,v}+\sigma_{I,w}}{\sigma_{I,z}+\sigma_{I,w}} = \frac{\beta\cdot\sigma_{I,z}+0+0}{\sigma_{I,z}+0} = \beta\cdot\frac{\sigma_{I,z}}{\sigma_{I,z}} = \beta$$

We should be aware that efficiency of this estimator could be low. Remaining errors w in the measurement q make no effect on consistency of estimator, but increase the variance of disturbance term and so increase standard errors of estimator.

(e) Why measurement errors in explanatory variable could be considered as a certain form of endogeneity in regression model? Describe in general terms a test procedure for determining whether the measurement error is significantly distorting the OLS estimates of the parameters.

Endogeneity means that the explanatory variable of the regression model is not exogenous and so its values at least partly determined inside the model, by the values of dependent variable and so disturbance term of the model. Measurement errors could be considered as a source of endogeneity as in the linear model they are part of both explanatory variable (which values are measured with error) and dependent variable, so they certainly correlate.

Durbin-Wu-Hausman test for endogeneity is appropriate here. The test statistic summarizes the differences between the OLS and IV estimates of the parameters and a chi-squared test is used to test whether this summary statistic is significant. If there is no measurement error, IV and OLS are both consistent but OLS is to be preferred because it is more efficient. If there is measurement error, IV is consistent but OLS is not.

There is also Hausman test for endogeneity (in the version proposed by Davidson and MacKinnon) includes two stages.

1) we regress suspect variable on all exogenous variables and instruments and retrieve the residuals (auxiliary regression).

2) then we re-estimate main regression including the residuals from the first (auxiliary) regression as additional regressors.

If the OLS estimates are consistent, then the coefficient on the first stage residuals should not be significantly different from zero.

So the significance of the residuals could be considered as a sign of possible endogeneity of suspect explanatory variable.

7. A researcher wishes to fit the following simple macroeconomic model:

$$C_t = \alpha + \beta Y_t + u_t \qquad (1)$$

$$I_t = \delta + \gamma r_t + v_t \qquad (2)$$

$$Y_t = C_t + I_t \qquad (3)$$

where C_t is aggregate consumption, I_t is aggregate investment, Y_t is aggregate output, r_t is the interest rate (assumed exogenous), and u_t, v_t are disturbance terms It may be assumed that u_t and v_t are distributed independently of each other and r_t. The researcher uses annual time series data for a certain country for 40 years, with C_t, I_t and Y_t measured in \$ billion at constant prices and the interest rate measured in percent per year. (Issues related to the non-stationary time series should be ignored.)

(a) The researcher wants to estimate all equations of the system above using OLS. Help him to perform estimation and explain why this is probably not the best solution (do not attempt here to explore consistency of the estimators as it will be the special task further). Explore identification of the system of equations above using both instrumental variable approach and order condition.

The dependent variable C_t includes the disturbance term of the first equation u_t, so through macroeconomic identity Y_t also becomes dependent of u_t, what violates Gauss-Markov conditions in the first equation. This is not applicable to the second equation as it has only exogenous explanatory variable in the right hand side of the equation. The second equation is certainly identified as OLS gives unbiased, efficient and thus consistent estimate of its parameters. The first equation is also identified as exogenous variable r_t could be used as an instrument for Y_t. The same could be derived from the order condition: the total number of endogenous variables and also equations is 3, so $G-1=2$. The first and the second equations missed two variables.

(b) Derive the reduced form system for the structural system of equations above and explain how ILS estimator could be obtained for the coefficient β of the first equation.

Substituting expression for I_t from the second equation into macroeconomic identity we get

$$Y_t = C_t + \delta + \gamma r_t + v_t$$

Now we can substitute the expression for Y_t from here into first equation:

$$C_t = \alpha + \beta(C_t + \delta + \gamma r_t + v_t)_t + u_t$$

or

$$C_t = \alpha + \beta\delta + \beta C_t + \beta\gamma r_t + \beta v_t + u_t$$

Grouping together homogeneous terms, and assuming $1-\beta$ non-zero we get

$$C_t = \frac{1}{1-\beta}[(\alpha + \beta\delta) + \beta\gamma r_t + (\beta v_t + u_t)]$$

or

$$C_t = \frac{\alpha + \beta\delta}{1-\beta} + \frac{\beta\gamma}{1-\beta} r_t + \frac{\beta v_t + u_t}{1-\beta}$$

The second equation do not need in transformation. The third equation could be obtained by substitution C_t from the obtained expression and I_t from the second equation:

$$Y_t = \frac{\alpha+\beta\delta}{1-\beta} + \frac{\beta\gamma}{1-\beta} r_t + \frac{\beta v_t + u_t}{1-\beta} + \delta + \gamma\, r_t + v_t$$

or

$$Y_t = \frac{\alpha+\beta\delta+\delta-\beta\delta}{1-\beta} + \frac{\beta\gamma+\gamma-\beta\gamma}{1-\beta} r_t + \frac{\beta v_t + u_t + v_t - \beta v_t}{1-\beta}$$

so the reduced form system is

$$C_t = \frac{\alpha+\beta\delta}{1-\beta} + \frac{\beta\gamma}{1-\beta} r_t + \frac{\beta v_t + u_t}{1-\beta} \qquad (1')$$

$$I_t = \delta + \gamma\, r_t + v_t \qquad (2')$$

$$Y_t = \frac{\alpha+\delta}{1-\beta} + \frac{\gamma}{1-\beta} r_t + \frac{u_t+v_t}{1-\beta} \qquad (3')$$

or simply

$$C_t = \pi_0 + \pi_1 r_t + w_t \qquad (1'')$$

$$I_t = \delta + \gamma\, r_t + v_t \qquad (2'')$$

$$Y_t = \pi_2 + \pi_3 r_t + z_t \qquad (3'')$$

where $\pi_1 = \frac{\beta\gamma}{1-\beta}$. From here $\pi_1 - \beta\pi_1 = \beta\gamma$ and $\pi_1 = \beta(\gamma+\pi_1)$ and so

$$\beta = \frac{\pi_1}{\gamma+\pi_1}$$

To get ILS estimator for β we need first estimate equation (2'') using OLS to get γ and then substitute it to expression (4).

$$\hat{\beta}^{ILS} = \frac{\hat{\pi}_1^{OLS}}{\hat{\gamma}^{OLS} + \hat{\pi}_1^{OLS}} \qquad (4)$$

Of course this estimator not being biased is consistent.

(c) Explain why inconsistent estimates would be obtained if OLS were used to fit the consumption function.

It is known that OLS estimator for β in the first equation is $\hat{\beta}^{OLS} = \frac{\mathrm{Cov}(C_t, Y_t)}{\mathrm{Var}(Y_t)}$ could be represented in the form $\hat{\beta}^{OLS} = \beta + \frac{\mathrm{Cov}(C_t, u_t)}{\mathrm{Var}(Y_t)}$. Now substitute expressions for C_t and Y_t from the reduced system

$C_t = \frac{1}{1-\beta}[(\alpha+\beta\delta) + \beta\gamma\, r_t + (\beta v_t + u_t)]$, $Y_t = \frac{1}{1-\beta}[(\alpha+\delta) + \gamma\, r_t + (u_t+v_t)]$

$$\hat{\beta}^{OLS} = \beta + \frac{\frac{1}{1-\beta}}{\frac{1}{(1-\beta)^2}} \cdot \frac{\mathrm{Cov}([(\alpha+\beta\delta)+\beta\gamma\, r_t + (\beta v_t + u_t)], u_t)}{\mathrm{Var}((\alpha+\delta)+\gamma\, r_t + (u_t+v_t))} =$$

$$\hat{\beta}^{OLS} = \beta + (1-\beta)\cdot \frac{\beta\gamma\,\mathrm{Cov}(r_t,u_t) + \beta^2\gamma\,\mathrm{Cov}(r_t,v_t) + \mathrm{Var}(u_t)}{\gamma^2\,\mathrm{Var}(r_t) + \mathrm{Var}(u_t) + \mathrm{Var}(v_t) + 2\gamma\,\mathrm{Cov}(r_t,u_t) + 2\gamma\,\mathrm{Cov}(r_t,v_t) + 2\,\mathrm{Cov}(u_t,v_t)}$$

Taking plim as $n \to \infty$ we get

$$\operatorname*{plim}_{n\to\infty}\hat{\beta}^{OLS} = \beta + (1-\beta)\cdot\frac{\beta\gamma\sigma_{r_t,u_t} + \beta^2\gamma\sigma_{r_t,v_t} + \sigma^2_{u_t}}{\gamma^2\sigma^2_{r_t} + \sigma^2_{u_t} + \sigma^2_{v_t} + 2\gamma\sigma_{r_t,u_t} + 2\gamma\sigma_{r_t,v_t} + 2\sigma_{u_t,v_t}}.$$

According to the assumptions of the model $\sigma_{r_t,u_t} = 0$, $\sigma_{r_t,v_t} = 0$ and $\sigma_{u_t,v_t} = 0$, so finally

$\operatorname*{plim}_{n\to\infty}\hat{\beta}^{OLS} = \beta + \dfrac{(1-\beta)\cdot\sigma^2_{u_t}}{\gamma^2\sigma^2_{r_t} + \sigma^2_{u_t} + \sigma^2_{v_t}}$ so the estimator is biased.

Since the data are time series, there is a possibility that $\operatorname{Var}(Y)$ might increase without limit, and hence that the bias term may vanish in large samples.

(d) Show that IV estimator of β based on using of r_t as an instrument is identical to the ILS estimator for this coefficient.

ILS estimator for β could be found

$$C_t = \frac{\alpha + \beta\delta}{1-\beta} + \frac{\beta\gamma}{1-\beta}r_t + \frac{\beta v_t + u_t}{1-\beta} \qquad (1')$$

Consider IV estimator of β in equation $C_t = \alpha + \beta Y_t + u_t$ (1), based on using of R as an instrument

$$\hat{\beta}^{IV} = \frac{\operatorname{Cov}(C_t, r_t)}{\operatorname{Cov}(Y_t, r_t)}.$$

Using macroeconomic identity $Y_t = C_t + I_t$ (3), this estimator becomes like this

$$\hat{\beta}^{IV} = \frac{\operatorname{Cov}(C_t, r_t)}{\operatorname{Cov}(Y_t, r_t)} = \frac{\operatorname{Cov}(C_t, r_t)}{\operatorname{Cov}(C_t, r_t) + \operatorname{Cov}(I_t, r_t)}.$$

Dividing both the numerator and the denominator by $\operatorname{Var}(r_t)$ we get

$$\hat{\beta}^{IV} = \frac{\dfrac{\operatorname{Cov}(C_t, r_t)}{\operatorname{Var}(r_t)}}{\dfrac{\operatorname{Cov}(C_t, r_t)}{\operatorname{Var}(r_t)} + \dfrac{\operatorname{Cov}(I_t, r_t)}{\operatorname{Var}(r_t)}}.$$

It could be easily seen that ratios in this expression are OLS estimators of the slopes of equations

$$C_t = \pi_0 + \pi_1 r_t + w_t \qquad (1'')$$

$$I_t = \delta + \gamma r_t + v_t \qquad (2'')$$

$$\hat{\pi}_1^{OLS} = \frac{\operatorname{Cov}(C_t, r_y)}{\operatorname{Var}(r_t)}, \qquad \hat{\gamma}^{OLS} = \frac{\operatorname{Cov}(I_t, r_y)}{\operatorname{Var}(r_t)}$$

So

$$\hat{\beta}^{IV} = \frac{\dfrac{\operatorname{Cov}(C_t, r_t)}{\operatorname{Var}(r_t)}}{\dfrac{\operatorname{Cov}(C_t, r_t)}{\operatorname{Var}(r_t)} + \dfrac{\operatorname{Cov}(I_t, r_t)}{\operatorname{Var}(r_t)}} = \frac{\hat{\pi}_1^{OLS}}{\hat{\pi}_1^{OLS} + \hat{\gamma}^{OLS}}$$

But this expression exactly the same as ILS estimator in (4)

$$\hat{\beta}^{ILS} = \frac{\hat{\pi}_1^{OLS}}{\hat{\gamma}^{OLS} + \hat{\pi}_1^{OLS}} \qquad (4)$$

So finally

$$\hat{\beta}^{IV} = \hat{\beta}^{ILS}$$

(e) Show that TSLS estimator of β based on variable r_t as an instrument is identical to the IV estimator for this coefficient based on the same instrument.

When using TSLS method for estimation coefficient β of equation

$$C_t = \alpha + \beta Y_t + u_t \quad (1)$$

using r_t as the only instrument we first at the stage 1 construct the following instrument

$$\hat{Y}_t = g_0 + g_1 r_t \qquad (5)$$

At the stage 2 we use this instrument to get consistent estimator of β:

$$\hat{\beta}^{TSLS} = \frac{\text{Cov}(C_t, \hat{Y}_t)}{\text{Cov}(Y_t, \hat{Y}_t)}$$

Substituting $\hat{Y}_t$ from (5) to the expression for TSLS estimator we get

$$\hat{\beta}^{TSLS} = \frac{\text{Cov}(C_t, \hat{Y}_t)}{\text{Cov}(Y_t, \hat{Y}_t)} = \frac{\text{Cov}(C_t, g_0 + g_1 r_t)}{\text{Cov}(Y_t, g_0 + g_1 r_t)} = \frac{g_1 \text{Cov}(C_t, r_t)}{g_1 \text{Cov}(Y_t, r_t)} = \frac{\text{Cov}(C_t, r_t)}{\text{Cov}(Y_t, r_t)} = \hat{\beta}^{IV}$$

So in the case of the exactly identified equation and only one instrument all three estimators are the same:

$$\hat{\beta}^{TSLS} = \hat{\beta}^{IV} = \hat{\beta}^{ILS}$$

The International College of Economics and Finance
Econometrics – 2015-2016.
Mid-year exam. 2015 December 24.
Suggested Solutions.

IMPORTANT: Start answering each question on the same page where the question is printed, then use additional pages from the same booklet. (ask for extra paper if necessary). Structure your answers in accordance with the structure of the questions. Testing hypotheses always state clearly null and alternative hypotheses provide critical value used for test, mentioning degrees of freedom and the significance level chosen for the test.

SECTION A

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

1. [15 marks]. We have hourly wage (measured in dollars), variable *W*, for male and female respondents. Dummy variable *male* is equal to 1 if respondent is male and 0 if respondent is female. Dummy variable *female* is equal to 1 if respondent is female and 0 if respondent is male. Consider three models

$$W_i = \alpha_1 + \alpha_2\, female_i + \varepsilon_i\,, \qquad i = 1, \ldots, n \quad (1),$$

$$W_i = \gamma_1 + \gamma_2 male_i + \delta_i\,, \qquad i = 1, \ldots, n \quad (2)$$

and

$$W_i = \beta_1 male_i + \beta_2\, female_i + u_i\,,\ i = 1, \ldots, n \quad (3)$$

(a) What are relations between the estimation results (coefficients and indicators of statistical quality) for two models (1) and (2)? Derive OLS estimators for the coefficients of two models (1) (α_1, α_2) and (3) (β_1, β_2). Show that $\hat{\alpha}_1 = \hat{\beta}_1$, $\hat{\alpha}_2 = \hat{\beta}_2 - \hat{\beta}_1$ and explain the intuition of these results.

Solution.

Estimate coefficients for model (1). Assume without any loss of generality that first m observations are male respondents and next f observations are female respondents $(m + f = n)$.
Thus:

$$\sum_{i=1}^{n}(W_i - \hat{\alpha}_1 - \hat{\alpha}_2\ female_i)^2 \to \min_{\hat{\alpha}_1, \hat{\alpha}_2} \qquad \sum_{i=1}^{n}(\mathrm{W}_i - \hat{\alpha}_1 - \hat{\alpha}_2\, female_i)^2 \to \min_{\hat{\alpha}_1, \hat{\alpha}_2}$$

$$\sum_{i=1}^{n}(W_i - \hat{\alpha}_1 - \hat{\alpha}_2\ female_i)^2 = \sum_{i=1}^{m}(W_i - \hat{\alpha}_1)^2 + \sum_{i=1}^{f}(W_i - \hat{\alpha}_1 - \hat{\alpha}_2)^2$$

We write the first order conditions (F.O.C.)

$$\begin{cases} -2\sum_{i=1}^{m}(W_i - \hat{\alpha}_1) - 2\sum_{i=1}^{f}(W_i - \hat{\alpha}_1 - \hat{\alpha}_2) = 0 \\ -2\sum_{i=1}^{f}(W_i - \hat{\alpha}_1 - \hat{\alpha}_2) = 0 \end{cases}$$

Substituting from the second equation into the first $\sum_{i=1}^{m}(W_i - \hat{\alpha}_1) = 0$ and $\hat{\alpha}_1 = \overline{W}_m$

From the second equation: $\sum_{i=1}^{f}(W_i - \hat{\alpha}_1 - \hat{\alpha}_2) = 0$, so $\sum_{i=1}^{f}(W_i - \overline{W}_m - \hat{\alpha}_2) = 0$ and $\hat{\alpha}_2 = \overline{W}_f - \overline{W}_m$

Second order conditions (S.O.C.) are trivial here.

Doing the same with the second equation (or just by analogy) we get

$\hat{\gamma}_1 = \overline{W}_f$ and $\hat{\gamma}_2 = \overline{W}_m - \overline{W}_f$

Comparing them with the previous results

$\hat{\alpha}_1 = \overline{W}_m$ and $\hat{\alpha}_2 = \overline{W}_f - \overline{W}_m$

we can see that $\hat{\gamma}_2 = -\hat{\alpha}_2$ and $\hat{\gamma}_1 = \hat{\alpha}_1 + \hat{\alpha}_2$

Intuition. We change the reference category, thus the sign of the slope coefficient must be changed and the intercept, which is, in fact, a wage level of the reference category, must be changed too.

Alternative solution. It is also possible to use theoretical equations to see the relationships between their coefficients.

$W_i = \alpha_1 + \alpha_2 female_i + \varepsilon_i = \alpha_1 + \alpha_2(1 - male_i) + \varepsilon_i = (\alpha_1 + \alpha_2) - \alpha_2 male_i + \varepsilon_i$

$W_i = \alpha_1 + \alpha_2\ female_i + \varepsilon_i = \alpha_1 + \alpha_2(1 - male_i) + \varepsilon_i = (\alpha_1 + \alpha_2) - \alpha_2 male_i + \varepsilon_i$

Thus models (1) and (2) are identical, R^2 are the same.

For the model (3)

$$\sum_{i=1}^{n}(W_i - \hat{\beta}_1 male_i - \hat{\beta}_2\ female_i)^2 \to \min_{\hat{\beta}_1, \hat{\beta}_2}$$

$$\sum_{i=1}^{n}(W_i - \hat{\beta}_1 male_i - \hat{\beta}_2\ female_i)^2 = \sum_{i=1}^{m}(W_i - \hat{\beta}_1)^2 + \sum_{i=1}^{f}(W_i - \hat{\beta}_2)^2$$

$$\hat{\beta}_1 = \frac{1}{m}\sum_{i=1}^{m} W_i = \overline{W}_m$$

$$\hat{\beta}_2 = \frac{1}{f}\sum_{i=1}^{f} W_i = \overline{W}_f$$

S.O.C. is trivial.

Thus: $\hat{\alpha}_1 = \hat{\beta}_1$, $\hat{\alpha}_2 = \hat{\beta}_2 - \hat{\beta}_1$.

Intuition. If we include two dummies instead of one dummy and intercept, we will not change the model. The estimator of the intercept is the average wage level for the reference categoty (male) in (1), the estimator of the slope coefficient shows the difference between average wage levels for two groups, because if respondent is female, than her predicted wage will be $\hat{\alpha}_1 + \hat{\alpha}_2$. In (3) the estimators of the both coefficients before dummies are the average wage levels of corresponding groups, because there are no intercept in the model, thus, for example, if respondent is male, than his predicted wage will be $\hat{\beta}_1$. The same restrictions hold for true parameters (see b) the second solution). The estimates of the coefficients are unbiased and consistent in this case, thus it is rather intuitivly, that these restrictions hold true for them too.

(b) Show that testing the hypothesis $\alpha_2 = 0$ for model (1), is equivalent to testing the hypothesis $\beta_1 = \beta_2$ for model (3).

We showed in a), that $\hat{\alpha}_2 = \hat{\beta}_2 - \hat{\beta}_1$

Thus $\mathrm{Var}(\hat{\alpha}_2) = \mathrm{Var}(\hat{\beta}_2 - \hat{\beta}_1)$, because data and formulas for righthand and lefthand parts are the same. Thus *t*-statistics will be the same (equal degrees of freedom, because we estimate two parameters in both cases).

Alternative solution. If we substitute $male_i = 1 - female_i$ into (3):

$$W_i = \beta_1(1 - female_i) + \beta_2 female_i + u_i = \beta_1 + (\beta_2 - \beta_1) female_i + u_i$$

Thus $\alpha_1 = \beta_1$, $\alpha_2 = \beta_2 - \beta_1$. Thus null and alternative hypotheses are the same in both cases, *t*-statistics are the same. So the tests under consideration are equivalent.

2. [15 marks]. New mobile phone company Tele2 is pursuing an aggressive sales strategy, offering customers the best plans. Tariff plan "Orange" implies the absence of a fixed fee. A rival company Beeline conducts sample study of Tele2 customers, and tries to estimate the regression

$$s_i = \beta\, t_i + u_i;\ i = 1, 2, ..., n, \quad (1)$$

where t_i is the number of minutes used, s_i - charge in roubles, t_i is supposed to be non-stochastic. Disturbance term u_i follows conventional conditions $E(u_i) = 0$, $E(u_i u_j) = 0$ if $i \neq j$ for all $i, j = 1, 2, ..., n$. As there is a significant variation of connection time between the observed customers, moreover, some of them use optional devices and the services of interenet, one can assume the presence of heteroscedasticity in the model of the type $E(u_i^2) = \sigma^2 t_i^2$.

(a) Let the assumption $E(u_i^2) = \sigma^2 t_i^2$ be valid. Describe the consequences of this for the OLS estimation of equation (1). What is WLS for estimation of equation (1)? Derive the weighted least squares (WLS) estimator $\hat{\beta}_{WLS}$, of β and show that it is unbiased.

Solution: The assumption $E(u_i^2) = \sigma^2 t_i^2$ describes one of the typical cases of heteroscedasticity, when standard deviation of the disturbance term is proportional to the 'proportionality factor' which is here variable t_i. Really $\operatorname{var}(u_i) = E(u_i^2) - (E(u_i))^2 = E(u_i^2) - 0 = E(u_i^2) = \sigma^2 t_i^2$, so $\sigma_{u_i} = \sigma t_i$

Under heteroscedasticity the estimator of β remains unbiased, but becomes inefficient, while tests are invalid. WLS is the method of applying OLS to the transformed equation where all variables, disturbance term (and constant if it is available in equation) are divided by the proportionality factor. As a result the variance of the disturnabce term becomes contant $\operatorname{var}(\frac{u_i}{t_i}) = \frac{\operatorname{var}(u_i)}{t_i^2} = \frac{E(u_i^2)}{t_i^2} = \frac{\sigma^2 t_i^2}{t_i^2} = \sigma^2$. To interpret the results we come back to the original equation multiplying estimated by WLS equation by the proportionality factor.

Here WLS is applying OLS to the weighted equation $\frac{s_i}{t_i} = \beta + \frac{u_i}{t_i}$.

Define $s_i^* = \frac{s_i}{t_i}$; $t_i^* = 1$; $u_i^* = \frac{u_i}{t_i}$. So WLS equation is $s_i^* = \beta t_i^* + u_i^*$. The regression without constant can be estimated by $\hat{\beta}_{OLS} = \frac{\sum s_i^* t_i^*}{\sum (t_i^*)^2} = \frac{\sum \frac{s_i}{t_i} \cdot 1}{\sum 1^2} = \frac{\sum \frac{s_i}{t_i}}{n}$

(b) Derive the variance of $\hat{\beta}_{WLS}$, and find out whether WLS estimator of β is consistent.

Solution: As $\hat{\beta}_{OLS} = \frac{1}{n}\sum \frac{s_i}{t_i} = \frac{1}{n}\sum \frac{\beta t_i + u_i}{t_i} = \beta + \frac{1}{n}\sum \frac{u_i}{t_i}$, it is not difficult to prove that the $\hat{\alpha}^{OLS}$ is unbiased:

$$E(\hat{\beta}_{OLS}) = E\left(\beta + \frac{1}{n}\sum \frac{u_i}{t_i}\right) = \beta + \frac{1}{n}\sum \frac{Eu_i}{t_i} = \beta$$

Now evaluate the variance of $\hat{\beta}_{OLS}$:

$$E(\hat{\beta}_{OLS}-\beta)^2 = E\left(\frac{1}{n}\sum\frac{u_i}{t_i}\right)^2 = \frac{1}{n}E\left(\sum\left(\frac{u_i}{t_i}\right)^2\right)+\frac{1}{n^2}E\left(\sum_{j\neq i}^{n}\frac{u_j u_i}{t_j t_i}\right)=\frac{1}{n^2}\sum\left(\frac{\sigma^2 t_i^2}{t_i^2}\right)=\frac{\sigma^2}{n}\xrightarrow[n\to\infty]{}0$$ so the sufficient condition for the consistency is satisfied.

3. [15 marks]. The 3d year student of ICEF is looking forward December mock exam in Econometrics, trying to anticipate her chance to pass the exam. For simplicity she assumes that all students have equal chance to pass the exam. Last year only 96 students from 171 passed December exam. Using simple proportion she estimates the probability p to pass December exam as $\hat{p}=\frac{96}{171}=0.56$. Upset by such a low probability she recalls that it was told in the course of Econometrics on the existence of such methods as OLS and MLE, so she has a hope that these methods will give her more optimistic outlook. Let Y be binary variable taking value 1 if the student passed the December exam (with probability p) and taking 0 for the student who did not pass the exam (with probability $1-p$), and let $y_1, y_2,..., y_n$ be the sample containing m values equal to 1, and $n-m$ values equal to zero.

(a) Using the sample above and assuming $y_i = p+u_i$ obtain OLS estimator $\hat{p}_{OLS}$ of probability p. What disadvantages of the linear probability model can be seen in this regression? Explain.

Solution: Let uss apply OLS method to get estimator $\hat{p}_{OLS}$ of p (we can not use ready made formulas as this equation contain only constant term). According equation $y_i = p+u_i$ predicted value of $\hat{y}_i$ for each student is p, so we have to minimize $S=\sum(y_i-\hat{y}_i)^2=\sum(y_i-p)^2\to\min$. Using necessary conditions for extremum we get $\frac{dS}{dp}=2\sum(y_i-p)=2(\sum y_i-np)=0$. So $\hat{p}=\frac{\sum y_i}{n}=\frac{m}{n}$. (Sufficient condition $\frac{d^2S}{dp^2}=2\frac{dS}{dp}\sum(y_i-p)=-2<0$ proves that we have maximum point).

Generally linear probability model has some shortcomings:

1) possible values outside the range $[0;1]$;

2) constant marginal effect;

3) heteroscedasticity;

4) the distibution of the disturbance term is far from normal distribution.

But in fact almost none of these shortcomings is really present in this model.

The estimated value of $\hat{p}=\frac{m}{n}$ is always in the range $[0;1]$. The second property (constant marginal effect) is not applicable here, as no explanatory variable is included in equation. Consider third property (heteroscedasticity). As $u_i = y_i - p$ so $\mathrm{var}(u_i)=\mathrm{var}(y_i-p)$ As y_i takes only two values 1 or 0, therefore u_i can take only two values $1-p$ when $y_i=1$ and p when $p_i=0$. Based on this we can write the probability distribution of u_i as:

y_i	u_i	$f(u_i)$
1	$1-p$	p
0	$-p$	$1-p$

This probability distribution also satisfies the assumption that:

$$\mathrm{E}(u_i)=(1-p)p+(-p)(1-p)=0$$

So we can write $\mathrm{var}(u_i)$ as:

$$\mathrm{var}(u_i)=\mathrm{E}(u_i^2)-(\mathrm{E}(u_i))^2=\mathrm{E}(u_i^2)-0=\mathrm{E}(u_i^2)=$$

$$= (1-p)^2 p + (-p)^2 (1-p) = p(1-p)(1-p+p) = p(1-p)$$; for all $i = 1, 2, ..., n$.

So the variance of the disturbance term is constant, so there is no heteroscedasticity.
The fourth property holds as $u_i = y_i - p$ takes only two values $1-p$ when $y_i = 1$ and p when $p_i = 0$, the distribution is obviously not normal.

(b) Using the same sample and assuming Bernoulli or binomial distribution derive MLE estimator $\hat{p}_{MLE}$ of p. Compare it with a naïve estimator $\hat{p} = \frac{m}{n}$ and OLS estimator $\hat{p}_{OLS}$. Were you able to please the student by increasing her chances to pass the exam?

Solution: To get $\hat{p}_{MLE}$ we have to maximize the likelihood function (approach based on Bernoulli distribution):
$L = P(y_1, y_2, ..., y_n) = P(y_1) \cdot P(y_2),..., P(y_n)$ (as observations in the sample are independent). Obviously $L = p^m (1-p)^{n-m}$, so $\ln L = m \ln p + (n-m)\ln(1-p)$. Using nesessary conditions for extremum we get $\frac{d(\ln L)}{dp} = \frac{m}{p} - \frac{(n-m)}{(1-p)} = 0 \Rightarrow m(1-p) = p(n-m) \Rightarrow \hat{p}_{MLE} = \frac{m}{n}$ (Sufficient condition $\frac{d^2(\ln L)}{dp^2} = -\frac{m}{p^2} - \frac{(n-m)}{(1-p)^2} < 0$ proves that this is maximum point). So we cannot please the student as all methods give identical results.

Note: An alternative approach is possible here: let X be binomial variable (number of successes in n independent trials with probability of success p). As we know $P(X = k) = C_n^k p^k (1-p)^{n-k}$
In our sample $k = n$, so likelihood function is $L = C_n^m p^m (1-p)^{n-m}$. The continuation of the story is the same as in Bernoulli approach).

SECTION B

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

4. [30 marks]. A student tries to estimate regression model of the type $y_t = \beta x_t + u_t\,; t = 1, 2, ..., T$, using OLS, but x_t is measured with error. Data is only available on x_t^*, where

$$x_t^* = x_t + v_t\,; t = 1, 2, ..., T$$

and $Eu_t = Ev_t = 0$, $E(u_t v_t) = E(x_t u_t) = E(x_t v_t) = 0$.

(a) If $\hat{\beta}$ is the ordinary least squares (OLS) estimator of β from regressing y_t on x_t^* without intercept, show that $\hat{\beta}$ is inconsistent and obtain the limiting expression for the large sample bias. The true value of the coefficient β may be assumed positive.

A student tries to estimate regression model of the type $y_t = \beta x_t + u_t\,; t = 1, 2, ..., T$, using OLS, but x_t is measured with error. Data is only available on x_t^*, where

$$x_t^* = x_t + v_t\,; t = 1, 2, ..., T$$

and $Eu_t = Ev_t = 0$, $E(u_t v_t) = E(x_t u_t) = E(x_t v_t) = 0$.

Solution:

(a) Direct proof of inconsistency: $y_t = \beta x_t + u_t$ where $x_t^* = x_t + v_t$

$$\hat{\beta} = \frac{\sum x_t^* y_t}{\sum x_t^{*2}} = \frac{\sum (x_t + v_t)(\beta x_t + u_t)}{\sum (x_t + v_t)^2} =$$

$$= \frac{\beta \sum x_t^2 + \sum x_t u_t + \beta \sum x_t v_t + \sum v_t u_t}{\sum x_t^2 + \sum v_t^2 + \sum x_t v_t}$$

$$\text{plim}\,\hat{\beta} = \frac{\text{plim}[\beta \sum x_t^2 + \sum x_t u_t + \beta \sum x_t v_t + \sum v_t u_t]/T}{\text{plim}[\sum x_t^2 + \sum v_t^2 + \sum x_t v_t]/T}$$

As it is known that $\text{plim}(\overline{X}) = \text{plim}\frac{1}{T}\sum x_t = E(X)$, so $\text{plim}\frac{1}{T}\sum x_t u_t = E(x_t u_t) = 0$,

$\text{plim}\frac{1}{T}\sum x_t v_t = E(x_t v_t) = 0$, $\text{plim}\frac{1}{T}\sum u_t v_t = E(u_t v_t) = 0$ and

$\text{plim}\frac{1}{\text{T}}\sum x_t^2 = E(x_t^2) = E(x_t^2) - 0 = E(x_t^2) - (E(x_t))^2 = \sigma_x^2$,

$\text{plim}\frac{1}{\text{T}}\sum v_t^2 = E(v_t^2) = E(v_t^2) - 0 = E(v_t^2) - (E(v_t))^2 = \sigma_v^2$

So

$$\text{plim}\,\hat{\beta} = \frac{\text{plim}\frac{1}{T}\sum x_t^2 + \text{plim}\frac{1}{T}\sum x_t u_t + \beta\,\text{plim}\frac{1}{T}\sum x_t v_t + \text{plim}\frac{1}{T}\sum v_t u_t}{\text{plim}\frac{1}{\text{T}}\sum x_t^2 + \text{plim}\frac{1}{\text{T}}\sum v_t^2 + \text{plim}\frac{1}{\text{T}}\sum x_t v_t}$$

or,

$$\text{plim}\,\hat{\beta} = \frac{\beta\sigma_x^2}{\sigma_x^2+\sigma_v^2} = \frac{\beta(\sigma_x^2+\sigma_v^2-\sigma_v^2)}{\sigma_x^2+\sigma_v^2} = \beta - \beta\frac{\sigma_v^2}{\sigma_x^2+\sigma_v^2} < \beta$$ (as β is assumed to be positive) $\Rightarrow$ Inconsistency.

(b) Suppose additionally that observable (without error) variable p_t is available, so that p_t correlates with x_t and $E(p_t u_t) = E(p_t v_t) = 0$. Suggest instrumental variable estimator for β based on p_t, and show that it is consistent. Discuss its advantages and disadvantages as to compare to conventional OLS estimator of β. The researcher by mistake uses variable p_t not as an instument for x_t but as a proxy for x_t, and uses OLS estimate for the slope as new estimate $\tilde{\beta}$ for β. Is the estimator $\tilde{\beta}_1$ consistent?

Solution: (b) Let us try to find $\frac{1}{T}\sum y_t p_t = \frac{1}{T}\sum(\beta x_t + u_t)p_t = \beta\frac{1}{T}\sum x_t p_t + \frac{1}{T}\sum u_t p_t$. As p_t does not correlates with u_t then for large samples $\frac{1}{T}\sum u_t p_t \to E(u_t p_t) = 0$

So for large samples $\frac{1}{T}\sum y_t p_t \approx \beta\frac{1}{T}\sum x_t p_t$ and it is naturally to use estimator $\hat{\beta}^{IV} = \dfrac{\frac{1}{T}\sum y_t p_t}{\frac{1}{T}\sum x_t p_t} = \dfrac{\sum y_t p_t}{\sum x_t p_t}$.

This estimator is consistent as going back

$$\text{plim}\,\hat{\beta}^{IV} = \text{plim}\frac{\sum y_t p_t}{\sum x_t p_t} = \text{plim}\frac{\sum(\beta x_t + u_t)p_t}{\sum x_t p_t} = \text{plim}\,\beta\frac{\sum x_t p_t + \sum u_t p_t}{\sum x_t p_t} = \beta + \text{plim}\frac{\sum u_t p_t}{\sum x_t p_t}$$

$$\text{plim}\,\hat{\beta}^{IV} = \beta + \frac{\text{plim}\frac{1}{T}\sum u_t p_t}{\text{plim}\frac{1}{T}\sum x_t p_t} = \beta$$, as $\text{plim}\frac{1}{T}\sum u_t p_t = E(u_t p_t) = 0$

If instrument p_t is used as proxy for x_t we have regression $y_t = \gamma p_t + s_t$ instead of $y_t = \beta x_t + u_t$. The estimator $\hat{\gamma}^{OLS}$ has nothing to do with β and is inconsistent for it. Let us show this. As p_t correlates with x_t we can assume that $x_t = \alpha p_t + r_t$. Substituting this expression into original regression we get

$y_t = \beta x_t + u_t = \beta(\alpha p_t + r_t) + u_t = \beta\alpha p_t + (u_t + \beta r_t)$. Of course $\widehat{\beta\alpha}$ does not tent to β. As an example one can take $\alpha = 2$: $x_t = 2p_t + r_t$.

(c) Let in addition to p_t the variable q_t (measured without error) is available, so that q_t correlates with x_t and $E(q_t u_t) = E(q_t v_t) = 0$. How TSLS can be used here to provide better estimator than IV with p_t or q_t?

Solution: (c) If two instruments are available p_t and q_t, then better instrument can be constructed as a linear combination of them $I_t = \delta_1 + \delta_2 p_t + \delta_3 q_t$. First step of TSLS – write down the regression $x_t = \delta_1 + \delta_2 p_t + \delta_3 q_t + \varepsilon_t$ (with conventional assumption on disturbance term ε_t. Using OLS obtain estimates of its coefficients $\hat{\delta}_1; \hat{\delta}_2; \hat{\delta}_3$, they are approximately most efficient ('best'). These estimators are used to evaluate estimated values of x_t:

$$\hat{x}_t = \hat{\delta}_1 + \hat{\delta}_2\, p_t + \hat{\delta}_3\, q_t$$

Of course $\hat{x}_t$ is also an instrument as it is likely to be correlated with x_t, and at the same time $E(\hat{x}_t u_t) = E[(\hat{\delta}_1 + \hat{\delta}_2 p_t + \hat{\delta}_3 q_t)u_t] = \hat{\delta}_1 E(u_t) + \hat{\delta}_2 E(u_t p_t) + \hat{\delta}_3 E(u_t q_t) = 0$ (the same for $E(I_t v_t)$.

Second step of TSLS is direct using of variable $\hat{x}_t$ in OLS estimation of the regression $y_t = \beta \hat{x}_t + u_t^*$.

Using additional information from both instruments TSLS allows to get optimal balance between them to provide better estimator than IV with p_t or q_t.

(d) In the above given model, suppose x_t was measured without error, y_t was measured with error and data was only available on y_t^* where

$$y_t^* = y_t + w_t \text{ and } E(w_t) = 0;\ E(u_t w_t) = E(x_t w_t) = 0 \text{ and } E(v_t w_t) = 0.$$

Let $\hat{\beta}$ be the OLS estimator of β from regressing y_t^* on x_t. Is $\hat{\beta}$ consistent? Explain in detail.

Solution:

(d)

$$\hat{\beta} = \frac{\sum x_t y_t^*}{\sum x_t^2} = \frac{\sum x_t (y_t + w_t)}{\sum x_t^2} = \frac{\sum x_t (\beta x_t + u_t + w_t)}{\sum x_t^2}$$
$$= \frac{\beta \sum x_t^2 + \sum x_t u_t + \sum x_t w_t}{\sum x_t^2}.$$

$$\text{plim}\, \hat{\beta} = \frac{\text{plim}[\beta \sum x_t^2 + \sum x_t u_t + \sum x_t w_t]/T}{\text{plim}[\sum x_t^2]/T}$$
$$= \frac{\beta \sigma_x^2}{\sigma_x^2} = \beta \Rightarrow \text{consistent.}$$

We use here the same consideration on connection between plims and expectations as in **(b).**

5. [30 marks]. The management of the publishing house, which publishes fashion magazine, is interested in the analysis of the factors affecting its financial state Y_t. The main factors are current publications P_t prepared by the staff of the magazine (additional publications allows to increase sales and improve financial state), and income from commercial ads A_t:

$$Y_t = \beta_1 P_t + \beta_2 A_t + u_t;\quad t = 1, 2, ..., T \qquad (1)$$

At the same time the volume of publications is determined by the financial state as preparation of publications about fashion requires a significant investment on photographers, models, transportation and location payments and so on.

$$P_t = \gamma_1 Y_t + v_t; \qquad t = 1, 2, ..., T \qquad (2)$$

The variable A_t may be supposed to be exogenous while P_t and Y_t are endogenous. All variables are taken in the index form. It may be supposed that stochastic variables $E(u_t) = E(v_t) = 0$, $E(u_t^2) = \sigma_u^2$, $E(v_t^2) = \sigma_v^2$, $E(A_t u_t) = 0$. $E(u_t v_t) = \sigma_{uv}$; $E(u_s v_t) = 0$ if $s \neq t$, for all $s, t = 1, 2, ..., T$ All slope coefficients are supposed to be positive.

(a) Explain why there is a problem of violation of the Gauss-Markov conditions in the system of econometric equations above. Derive the reduced form equations for the system above and use it to explore the problem of identification of both equations.

Solution:

(a) Two endogenous variables Y_t and P_t in equations (1) and (2) constitute circular dependence $eq.(1): P_t \to Y_t$ $eq.(2): Y_t \to P_t$ and so there is possibility that disturbance term u_t which is a part of Y_t in eq(1) will correlate with P_t $(eq.(2): Y_t \to P_t)$ and thus P_t being explanatory variable of eq.(1) will correlate with disturbance term u_t of the same equation (possible violation of the GM conditions). The same with disturbance term v_t of eq.(2) that can correlate with explanatory variable Y_t of the same equation.

To understand that this possibility comes into reality we have to derive reduced form equations and using them to investigate large sample bias when OLS is applied to estimation of initial structured form equations (see further **(a)** and **(b)**.

To obtain reduces form equation we substitute expressions for endogenous variables from eq.(1) into eq.(2) and vice versa.

$$Y_t = \beta_1 P_t + \beta_2 A_t + u_t = \beta_1(\gamma_1 Y_t + v_t) + \beta_2 A_t + u_t = \beta_1\gamma_1 Y_t + \beta_2 A_t + u_t + \beta_1 v_t\,,$$

so $(1-\beta_1\gamma_1)Y_t = \beta_2 A_t + u_t + \beta_1 v_t$ and thus assuming $1-\beta_1\gamma_1 \neq 0$, we get

$$Y_t = \frac{\beta_2}{1-\beta_1\gamma_1} A_t + \frac{u_t + \beta_1 v_t}{1-\beta_1\gamma_1}$$

By analogy

$$P_t = \gamma_1 Y_t + v_t = \gamma_1(\beta_1 P_t + \beta_2 A_t + u_t) + v_t = \beta_1\gamma_1 P_t + \gamma_1\beta_2 A_t + v_t + \gamma_1 u_t\,,$$

So $(1-\beta_1\gamma_1)P_t = \gamma_1\beta_2 A_t + v_t + \gamma_1 u$ and under the same assumption

$$P_t = \frac{\gamma_1\beta_2}{1-\beta_1\gamma_1} A_t + \frac{v_t + \gamma_1 u}{1-\beta_1\gamma_1}$$

Using notations

$$\frac{\beta_2}{1-\beta_1\gamma_1}=\pi_1 \quad (3), \qquad \frac{u_t+\beta_1 v_t}{1-\beta_1\gamma_1}=u_t^*$$

$$\frac{\gamma_1\beta_2}{1-\beta_1\gamma_1}=\pi_2 \quad (4), \qquad \frac{v_t+\gamma_1 u}{1-\beta_1\gamma_1}=v_t^*$$

we can write reduced form system in the short form

$$Y_t=\pi_1 A_t+u_t^*$$
$$P_t=\pi_2 A_t+v_t^*$$

All equations of this system include only exogenous variable A_t in their right hand side so they can be estimated using OLS

$$\hat{\pi}_1=\frac{\sum Y_t A_t}{\sum A_t^2};\ \hat{\pi}_2=\frac{\sum P_t A_t}{\sum A_t^2}$$

As from (3) and (4) $\gamma_1=\dfrac{\frac{\gamma_1\beta_2}{1-\beta_1\gamma_1}}{\frac{\beta_2}{1-\beta_1\gamma_1}}=\dfrac{\pi_2}{\pi_1}$ we can use this to get ILS estimator for $\hat{\gamma}_1^{ILS}=\dfrac{\hat{\pi}_2}{\hat{\pi}_1}$.

It is definitely consistent as $\operatorname{plim}\hat{\gamma}_1^{ILS}=\dfrac{\operatorname{plim}\hat{\pi}_2}{\operatorname{plim}\hat{\pi}_1}=\dfrac{\pi_2}{\pi_1}=\gamma_1$, so eq.(2) is exactly identified (for all its parameters consistent estimators can be found).

On the other hand the system of equations (3), (4) includes three unknowns $\beta_1, \beta_2, \gamma_1$ and only two equations so it cannot have unique solutions for all $\beta_1, \beta_2, \gamma_1$, so consistent estimators for equation (1) cannot be found and thus this eqiation is underidentified.

(b) Investigate the sign of the bias in the slope coefficient if OLS is used to fit equation (2).

Solution:

(b) Let equation $P_t=\gamma_1 Y_t+v_t;$ (2) is estimated using OLS, $\hat{\gamma}_1=\dfrac{\sum Y_t P_t}{\sum Y_t^2}$. First we use structural equation $P_t=\gamma_1 Y_t+v_t;$ (2) to get decomposition into fixed and random components

$$\hat{\gamma}_1=\frac{\sum Y_t P_t}{\sum Y_t^2}=\frac{\sum Y_t(\gamma_1 Y_t+v_t)}{\sum Y_t^2}=\gamma_1\frac{\sum Y_t^2}{\sum Y_t^2}+\frac{\sum Y_t v_t}{\sum Y_t^2}=\gamma_1+\frac{\sum Y_t v_t}{\sum Y_t^2}$$

To investigate properties of this estimator now we use expression for Y_t from reduced form equation $Y_t=\dfrac{1}{1-\beta_1\gamma_1}(\beta_2 A_t+u_t+\beta_1 v_t)$.

$$\hat{\gamma}_1=\gamma_1+\frac{\sum Y_t v_t}{\sum Y_t^2}=\gamma_1+\frac{\frac{1}{1-\beta_1\gamma_1}\sum(\beta_2 A_t+u_t+\beta_1 v_t)v_t}{\frac{1}{(1-\beta_1\gamma_1)^2}\sum(\beta_2 A_t+u_t+\beta_1 v_t)^2}=\gamma_1+(1-\beta_1\gamma_1)\frac{\beta_2\sum A_t v_t+\sum u_t v_t+\beta_1\sum v_t^2}{\sum(\beta_2 A_t+u_t+\beta_1 v_t)^2}=$$

$$\hat{\gamma}_1=\gamma_1+(1-\beta_1\gamma_1)\frac{\frac{1}{T}(\beta_2\sum A_t v_t+\sum u_t v_t+\beta_1\sum v_t^2)}{\frac{1}{T}(\beta_2^2\sum A_t^2+\sum u_t^2+\beta_1^2\sum v_t^2+2\beta_2\sum A_t u_t+2\beta_2\beta_1\sum A_t v_t+2\beta_1\sum u_t v_t)}$$

Now $\operatorname{plim}\frac{1}{T}\sum A_t v_t=E(A_t v_t)=0$, $\operatorname{plim}\frac{1}{T}\sum A_t u_t=E(A_t u_t)=0$, $\operatorname{plim}\frac{1}{T}\sum u_t v_t=E(u_t v_t)=0$,

$\text{plim}\frac{1}{T}\sum v_t^2=\sigma_v^2$, $\text{plim}\frac{1}{T}\sum u_t^2=\sigma_u^2$, $\text{plim}\frac{1}{T}\sum A_t^2=\sigma_A^2$. So $\text{plim}\hat{\gamma}_1=\gamma_1+(1-\beta_1\gamma_1)\beta_1\frac{\sigma_v^2}{\beta_2^2\sigma_A^2+\sigma_u^2+\beta_1^2\sigma_v^2}$

From the meaning of reduced equation $Y_t=\frac{1}{1-\beta_1\gamma_1}(\beta_2A_t+u_t+\beta_1v_t)$ the expression $1-\beta_1\gamma_1$ is likely to be positive, we are told that all slope coefficients are positive, so we can say that OLS overestimates the value of γ_1.

(c) Suppose we have additionally data on F_t - miscellaneous publications prepared by freelancers (F_t is supposed to be exogenous, correlated with Y_t and not correlated with v_t. How this variable can be used to construct instrumental variable (IV) estimator $\hat{\gamma}_1^{IV}$ for γ_1? Is this IV estimator consistent? The researcher by mistake uses variable F_t not as an instument for Y_t but as a proxy for Y_t, and uses OLS estimate for the slope as new estimate $\tilde{\gamma}_1$ for γ_1. Is the estimator $\tilde{\gamma}_1$ consistent?

Solution:

(c) Let us try to find $\frac{1}{T}\sum P_tF_t=\frac{1}{T}\sum(\gamma_1Y_t+v_t)F_t=\beta\frac{1}{T}\sum Y_tF_t+\frac{1}{T}\sum v_tF_t$. As F_t does not correlates with v_t then for large samples $\frac{1}{T}\sum v_tF_t\to E(v_tF_t)=0$

So for large samples $\frac{1}{T}\sum P_tF_t\approx\gamma_1\frac{1}{T}\sum Y_tF_t$ and it is naturally to use estimator $\hat{\gamma}_1^{IV}=\frac{\sum P_tF_t}{\sum Y_tF_t}$.

This estimator is consistent as

$$\text{plim}\,\hat{\gamma}_1^{IV}=\text{plim}\frac{\sum P_tF_t}{\sum Y_tF_t}=\text{plim}\frac{\sum(\gamma_1Y_t+v_t)F_t}{\sum Y_tF_t}=\text{plim}\,\gamma_1\frac{\sum Y_tF_t+\sum v_tF_t}{\sum Y_tF_t}=\gamma_1+\text{plim}\frac{\sum v_tF_t}{\sum Y_tF_t}$$

$$\text{plim}\,\hat{\gamma}_1^{IV}=\gamma_1+\frac{\text{plim}\frac{1}{T}\sum v_tF_t}{\text{plim}\frac{1}{T}\sum Y_tF_t}=\gamma_1,\ \text{as}\ \text{plim}\frac{1}{T}\sum v_tF_t=E(v_tF_t)=0$$

Here $E(v_tF_t)=E(v_tF_t)-E(v_t)E(F_t)=\text{cov}(v_t;F_t)$ as $E(v_t)=0$

If instrument F_t is used as proxy for Y_t we have regression $P_t=\delta_1F_t+s_t$ instead of $P_t=\gamma_1Y_t+v_t$. The estimator $\hat{\delta}_1^{OLS}$ has nothing to do with γ_1 and is inconsistent for it. Let us show this. As F_t correlates with Y_t we can assume that $Y_t=\alpha F_t+r_t$. Substituting this expression into original regression we get $P_t=\gamma_1Y_t+v_t=\gamma_1(\alpha F_t+r_t)+v_t=\gamma_1\alpha F_t+(v_t+\gamma_1r_t)$. Of course $\hat{\gamma_1\alpha}$ does not tent to γ_1. As an example one can take $\alpha=2$: $Y_t=2F_t+r_t$. The estimator $2\hat{\gamma}_1$ will be approximately twice as much as γ_1.

(d) Use order condition to investigate the identification of equations (1) and (2). How to transform equations using new variable F_t to get two exactly identified equations?

By mistake the researcher includes new variable F_t into first equation

$$Y_t = \beta_1 P_t + \beta_2 A_t + \beta_3 F_t + s_t \ ; \quad t = 1, 2, ..., T \quad (3)$$
$$P_t = \gamma_1 Y_t + r_t \ ; \quad t = 1, 2, ..., T \quad (4)$$

How this changes the situation with identification of both equation? Use both order condition and the instrumental variables approach to answer this question.

Solution: **(d)** Consider equations (1) and (2)

$$Y_t = \beta_1 P_t + \beta_2 A_t + u_t \ ; \quad (1)$$
$$P_t = \gamma_1 Y_t + v_t ; \quad (2)$$

The number of equation here is $G = 2$, $G - 1 = 1$ and there is 1 missed variable from (2) and no one from (1). So equation (2) is exactly identified while equation (1) is underidentified.

Instrumental variable approach gives the same: there is one available instrument A_t for endogenous variable Y_t in the right hand side of the equation (2) but there is no available instrument for P_t in equation (1).

The situation changes in equations (3)-(4).

$$Y_t = \beta_1 P_t + \beta_2 A_t + \beta_3 F_t + s_t \ ; \quad (3)$$
$$P_t = \gamma_1 Y_t + r_t \ ; \quad (4)$$

Now $2 > G - 1 = 1$ variables are missed from equation (2) so it is overidentified, while equation (1) remains underidentified.

The International College of Economics and Finance
Econometrics – 2016-2017.
Mid-year exam. 2016 December 29.

SECTION A

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

Question 1

1. [15 marks]. A researcher is studying the dependence of the winter sports equipment (skies, snowboards etc.) expenditures (variable $WINTER$) on the value of personal income PI_t, the value of taxes TAX_t, (all these three variables are measured in billions of US dollars), and relative price index of sports equipment $PREWINTER_t$ (in % of winter sports equipment prices to prices of total personal expenditures) for US economy in 2000-2016. (All questions assotiated with non-stationary time series should be ignored here).

$$\underset{}{\widehat{WINTER_t}} = \underset{(2.60)}{6.35} + \underset{(0.00173)}{0.0123} \cdot PI_t - \underset{(0.00652)}{0.0132} TAX_t - \underset{(0.0387)}{0.108} \cdot PREWINTER_t \qquad \begin{array}{l} R^2 = 0.9707 \\ RSS = 0.9187 \end{array} \qquad \textbf{(eq.1)}$$

(a) [7 marks]. Give an interpretation to the coefficient of variable TAX_t. Test the significance of coefficients of each variable and the significance of the equation as a whole. How your conclusion changes if it can be assumed that the coefficients of the variables TAX_t and $PREWINTER_t$ cannot be positive and coefficient of PI_t cannot be negative .

(a1) If taxes increases by 1 billion of dollars the expenditures on sports equipment decrease by 13.2 millions of dollars keeping personal income and relative prices of sports equipment constant.

(a2) Let theoretical model be

$$WINTER_t = \beta_1 + \beta_2 \cdot PI_t + \beta_3 \cdot TAX_t + \beta_4 \cdot WINTER_t + u_t$$

To test the significance of the coefficients $\begin{pmatrix} H_0 : \beta = 0, \\ H_a : \beta \neq 0 \end{pmatrix}$ let us evaluate *t*-statistics for the coefficients:

$t_{PI} = \dfrac{0.0123}{0.00173} = 7.11$, $t_{TAX} = -2.03$, $t_{PRESPORT} = -2.79$ ($df = n - k = 17 - 4 = 13$),

while $t^{crit}_{5\%}(df = 13) = 2.160$, $t^{crit}_{1\%}(df = 13) = 3.012$, so coefficient of PI_t is significant at 1% level, coefficient of $PREWINTER_t$ is significant only at 5% level, and coefficient of TAX_t is insignificant.

(a3) To test the significance of all coefficients simultaneously $\begin{pmatrix} H_0 : \beta_2 = \beta_3 = \beta_4, \\ H_a : \text{otherwise} \end{pmatrix}$ one should evaluate F-statistics

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0.9707/3}{(1-0.9707)/13} = 143.56$$

$F^{crit}_{1\%}(3,13) = 5.74$, so H_0 is rejected.

(a4) Using one-sided tests $\begin{pmatrix} H_0 : \beta = 0, \\ H_a : \beta < 0 \end{pmatrix}$ or $\begin{pmatrix} H_0 : \beta = 0, \\ H_a : \beta > 0 \end{pmatrix}$ could help in some situations to make significant the coefficient that is insignificant when using two-sided test.

In our situation $\begin{pmatrix} H_0 : \beta = 0, \\ H_a : \beta < 0 \end{pmatrix}$ $t^{crit}_{5\%}(\text{one}-\text{side}, 13) = 1.771$, $t^{crit}_{1\%}(\text{one}-\text{side}, 13) = 2.65$, so coefficient of TAX_t becomes significant and coefficient of $PREWINTER_t$ becomes significant even at 1% level. Coefficient of PI_t does not require one sided testing.

(b) [8 marks]. One of the colleagues of the researcher advised her to use variable DPI_t (disposable personal income) instead of PI_t (by definition $DPI_t = PI_t - TAX_t$), the corresponding equation is (eq.2)

$$\hat{WINTER}_t = \underset{(2.60)}{6.35} + \underset{(0.00173)}{0.0123} \cdot DPI_t - \underset{(0.00517)}{0.000953} \cdot TAX_t - \underset{(0.0387)}{0.108} \cdot PREWINTER_t \qquad R^2 = 0.9707, \quad RSS = 0.9187 \qquad \textbf{(eq.2)}$$

Another colleague advised the researcher to drop variable TAX_t from equations with the results:

$$\hat{WINTER}_t = \underset{(2.46)}{6.44} + \underset{(0.00117)}{0.0121} \cdot DPI_t - \underset{(0.0344)}{0.108} \cdot PREWINTER_t \qquad R^2 = 0.9706, \quad RSS = 0.9211 \qquad \textbf{(eq.3)}$$

Explain why the coefficients and their standard errors for variables PI_t and DPI_t and for variable $PREWINTER_t$, as well as R^2 and RSS are exactly the same in equations (eq1) and (eq2), while coefficient of variable TAX_t in equation (eq.2) is different from that in equation (eq.1) .
The researcher realized that equation (eq.3) is a restricted version of the equation (eq.1). What is the restriction? Is it rejected?
What equation from eq1-eq2-eq3 would you recommend for further analysis and why?

(b1) The theoretical equation corresponding (eq2) is

$$WINTER_t = \alpha_1 + \alpha_2 \cdot DPI_t + \alpha_3 \cdot TAX_t + \alpha_4 \cdot PREWINTER_t + u_t$$

Substituting $DPI_t = PI_t - TAX_t$ into this equation we get

$$WINTER_t = \alpha_1 + \alpha_2 \cdot PI_t - \alpha_2 \cdot TAX_t + \alpha_3 \cdot TAX_t + \alpha_4 \cdot PREWINTER_t + u_t$$

or

$$WINTER_t = \alpha_1 + \alpha_2 \cdot PI_t + (\alpha_3 - \alpha_2) \cdot TAX_t + \alpha_4 \cdot PREWINTER_t + u_t$$

what is identical to

$$WINTER_t = \beta_1 + \beta_2 \cdot PI_t + \beta_3 \cdot TAX_t + \beta_4 \cdot PREWINTER_t + u_t$$

with the same variables and the same disturbance term. So all characteristics of this equation – estimates of coefficients, their standard errors (except those related to the variable TAX_t), R^2 and RSS should be equal for both equations.

There is alternative and simpler way of explanation. It is obvious that both equation express the relationship of $WINTER_t$ from the same set of variables PI_t , TAX_t and $PREWINTER_t$. So R^2 and RSS , the coefiicient of $PREWINTER_t$ should be the same. As for coefficients of different variables PI_t and DPI_t they also should be equal, as interpretation of the coefficients of the multiple regression model requires that all other variables should remain constant: if PI_t increases by 1 dollar keeping TAX_t constant the variable DPI_t also increases by 1 dollar from equality $DPI_t = PI_t - TAX_t$ so their marginal effect to $WINTER_t$ should be the same.

(b2) The coefficient of TAX_t in equation

$$WINTER_t = \alpha_1 + \alpha_2 \cdot DPI_t + \alpha_3 \cdot TAX_t + \alpha_4 \cdot PREWINTER_t + u_t$$

is less in absolute value than in equation

$$WINTER_t = \alpha_1 + \alpha_2 \cdot PI_t + (\alpha_3 - \alpha_2) \cdot TAX_t + \alpha_4 \cdot PREWINTER_t + u_t$$

derived in (b1), taking into account that α_3 should be negative from economic consideration (and $-\alpha_2$ is also negative). Being smaller it can easily become insignificant even in one sided test. Additional factor should be mentioned: relatively bigger standard errors caused by multicollinearity (as taxes are already included into equation through disposable personal income) also lead to insignificance.

There is alternative way of explanation. The interpretation of the coefficients of the multiple regression model requires that all other variables should remain constant. In interpretation of equation

$$WINTER_t = \alpha_1 + \alpha_2 \cdot DPI_t + \alpha_3 \cdot TAX_t + \alpha_4 \cdot PREWINTER_t + u_t$$

if TAX_t increases by 1 dollar keeping DPI_t constant the variable PI_t will also increase from the equality $PI_t = DPI_t + TAX_t$. So the coefficient of TAX_t shows combined effect of simultaneous increase of TAX_t and PI_t by the same value, the result of that is economically uncertain, what is clearly explains the insignificance of this coefficient.

(b3) The equation

$$WINTER_t = \gamma_1 + \gamma_2 \cdot DPI_t + \gamma_4 \cdot PREWINTER_t + u_t$$

is the restricted version of the equation

$$WINTER_t = \alpha_1 + \alpha_2 \cdot PI_t - \alpha_2 \cdot TAX_t + \alpha_3 \cdot TAX_t + \alpha_4 \cdot PREWINTER_t + u_t$$

under restriction $\alpha_3 = -\alpha_2$. It is not rejected as under $\begin{pmatrix} H_0 : no\ difference\ between\ equations, \\ H_a : otherwise \end{pmatrix}$ the F-statistics is $F = \frac{9211 - 9187}{9187} \cdot 13 = 0.034$ what is definitely insignificant ($F^{crit}_{5\%}(1,13) = 4.67$), so H_0 is not rejected, and so simplier equation (eq.3) should be chosen (more efficient estimates, less standard errors).

Question 2

2. [15 marks]. Two systems of econometric equations are considered

(A)
(1) $y_{1t} = \alpha_0 + \alpha_1 y_{2t} + \alpha_2 x_{1t} + \varepsilon_{1t}$
(2) $y_{2t} = \beta_0 + \beta_1 y_{1t} + \varepsilon_{2t}$,

(B)
(1) $y_{1t} = \alpha_0 + \alpha_1 y_{2t} + \alpha_2 y_{3t} + \alpha_3 x_{1t} + \varepsilon_{1t}$
(2) $y_{2t} = \beta_0 + \beta_1 y_{1t} + \varepsilon_{2t}$,
(3) $y_{3t} = \gamma_0 + \gamma_1 x_{1t} + \gamma_2 x_{2t} + \gamma_3 x_{3t} + \varepsilon_{3t}$

where the variables y_{1t}, y_{2t}, y_{3t} are endogenous, and x_{1t}, x_{2t}, x_{3t} are exogenous. All ε_t are distributed independently of each other and independently of all variables of the system.

(a) [8 marks]. Examine the identifiability of each equation of the system **(A)**. How does the situation change for the system **(B)**?
You may use in your analysis the order condition or the rule based on the choice of instrumental variable.
Show that in general case order condition is equivalent to the rule based on the choice of instrumental variable.
(Use standard notations: G – number of equations (endogenous variables in the system), R – number of restrictions (number of variables excluded from equation), j – the number of endogenous variables excluded from the equation under consideration).

(a) The theoretical equation corresponding (eq2) is

Order condition: $R \geq G - 1$ where R is the number of restrictions imposed on the equation. G is the number of endogenous variables in the complete model which is also equal to the number of equations.

$R > G - 1 \quad \Rightarrow$ overidentified
$R = G - 1 \quad \Rightarrow$ exactly identified
$R < G - 1 \quad \Rightarrow$ underidentified.

Applying order condition to the system **(A)**:
First equation $R = 0$, $G - 1 = 2 - 1 = 1$. As $R < G - 1$, equation is underidentified.
Second equation $R = 1$, $G - 1 = 2 - 1 = 1 \quad \Rightarrow$ exactly identified.

For the system **(B)** situation changes:
First equation $R = 2$, $G - 1 = 3 - 1 = 2 \quad \Rightarrow \quad R = G - 1$, equation is exactly identified.
Second equation $R = 4$, $G - 1 = 3 - 1 = 2$. As $R > G - 1$, equation is overidentified.

Third equation: there is no problem with identification. The equation has only exogenous variables on the right hand side. OLS can be applied.

Alternative approach, based on instruments.
Instrumental variable is an exogenous variable that is not present in equation on its own right.
For the system **(A)**:
First equation has one endogenous variable y_{2t} that is needed to be instrumented, but the only exogenous variable is present in equation. Using it as an instrument would cause perfect multicollinearity. Underidentified.
Second equation has one endogenous variable y_{1t} that is needed to be instrumented, and there is available instrument x_{1t}; exactly identified.

For the system **(B)**:
First equation has two endogenous variables y_{2t} and y_{3t} that is needed to be instrumented, and two possible instruments x_{2t} and x_{3t} $\Rightarrow$ exactly identified.
Second equation has one endogenous variable y_{2t} that is needed to be instrumented, and there are three available instruments x_{1t}, x_{2t}, x_{3t} $\Rightarrow$ overidentified.
Third equation: no problem with identification.

Equivalence of two rules of identification.
According to the rule based on the choice of instruments, an equation is identified if the number of available instruments is not less than the number of endogenous variables on the right hand side of the equation.
Suppose we have an equation that cannot be estimated applying OLS due to the presence of endogenous variables on the right hand side. Let G – number of equations (endogenous variables in the system), and R – number of variables excluded from the equation under consideration (number of restrictions). Suppose that j endogenous variables are missing from the equation. Then $G-j-1$ are available on the right side, and at least $G-j-1$ instruments are needed. So the minimum number of variables missing from the equation is $j+(G-j-1)=G-1$. So the order condition is satisfied.
Now let order condition for identification holds: equation is likely to be identified if $G-1$ or more variables are missing from it. Suppose again that j endogenous variables are missing from the equation, so $G-j-1$ endogenous variables are available on the right side of the equation. But if we count the number of missing exogenous variables (available instruments) it would be the same: $G-j-1$ ($G-1$ total missing variables minus j endogenous variables missing), so the number of available instruments is not less than the number of endogenous variables on the right hand side of the equation. It means that two rules are equivalent.

(b) [7 marks]. How would you estimate each equation of the systems **(A)** and **(B)**. What are properties of estimators obtaining by applying different methods to the equations of each system?
(It is not expected here students to derive reduced form equations).

(b) OLS estimates of both equations of the system **(A)** are inconsistent. For underidentified equation (1) there is no method to get consistent estimates. For exactly identified equation (2) all three methods (ILS, IV and TSLS) allows to get the same consistent estimates.
In system **(B)** only equation (3) OLS gives BLUE estimates (under corresponding conditions). For equation (1) it is possible to use ILS, IV or TSLS and to get consistent estimates. Equation (2) can be consistently estimated only by TSLS.

Question 3

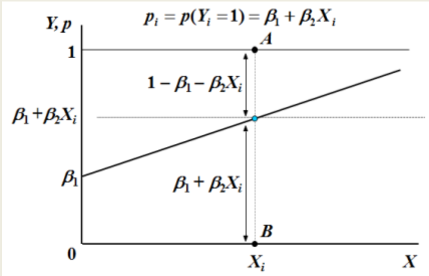
3. [15 marks]. Consider a linear probability model

$$Y_i = \beta_1 + \beta_2 X_i + u_i\,;\; i = 1, 2, ..., n\,,$$

where $Y_i = 1$ if the event takes place, $Y_i = 0$ otherwise and $E(u_i) = 0$.

(a) [8 marks]. Explain why there is a problem of heteroscedasticity by evaluating the variance of the disturbance term and examining its properties. What other problems are connected with linear probability model?

Solution



a) [8 marks]. Model is: $Y_i = \beta_1 + \beta_2 X_i + u_i$, $i = 1, 2, ..., n$

$Y_i = 1$ if the event takes place, $Y_i = 0$ otherwise.

We can interpret $\beta_1 + \beta_2 X_i$ as the probability that the event will occur, given X_i. then the probability of the event will occur, given X_i. is $1-\beta_1-\beta_2 X_i$.

From here we get $E[Y_i \mid X_i] = 1\cdot(\beta_1+\beta_2X_i) + 0\cdot(1-\beta_1-\beta_2X_i) = \beta_1+\beta_2X_i$,

As Y_i takes only two values 1 or 0, therefore u_i can take only two values: $1-\beta_1-\beta_2X_i$ when $Y_i = 1$ and $-\beta_1-\beta_2X_i$ when $Y_i = 0$. Based on this we can write the probability distribution of u_i as

Y_i	u_i	$f(u_i)$
1	$1-\beta_1-\beta_2X_i$	$\beta_1+\beta_2X_i$
0	$-\beta_1-\beta_2X_i$	$1-\beta_1-\beta_2X_i$

As $E(u_i) = 0$ we can write $V(u_i)$ as

$$V(u_i) = E[u_i^2] = (1-\beta_1-\beta_2X_i)^2(\beta_1+\beta_2X_i) + (-\beta_1-\beta_2X_i)^2(1-\beta_1-\beta_2X_i) =$$
$$= (1-\beta_1-\beta_2X_i)(\beta_1+\beta_2X_i)\times[(1-\beta_1-\beta_2X_i)+(\beta_1+\beta_2X_i)] =$$
$$= (\beta_1+\beta_2X_i)(1-\beta_1-\beta_2X_i)\quad \forall i = 1, 2, ..., n$$

Hence the disturbance term is heteroscedastic ($V(u_i)$ is different for different points of the sample). This will make *OLS* estimators inefficient. Weighted least squares can be used to obtain efficient estimators of β_1 and β_2 .

Additional problems

i. In many cases the estimated probability $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ will be negative or greater than 1.

ii. As the distribution of the disturbance term only takes two values, it is not continuous. Since *u* does not have a normal distribution, the standard errors and test statistics are invalid. This implies that usual test statistics are invalid.

iii. Marginal effect of the factor X_i is constant (equal to β_2).

(b) [7 marks]. The alternative approach to estimation of linear regression is based on the using of so called logit model estimated by the maximum likelihood method. Explain the concept of the likelihood function and the maximum likelihood estimator, and briefly describe its properties.

Derive the expression for the loglikelihood function for the logit model.

(It is not expected here the students to solve maximization problem)

Solution and marking

The maximum likelihood estimator (MLE) is an estimator which maximizes the likelihood function. Likelihood function gives the joint probability (or probability density function) as a function of the parameter, given the sample observations. It is widely used in econometrics for estimating regression parameters alternatively to OLS estimation.

Suppose the population X is distributed following the probability density function $f(x;\theta)$ with unknown parameter θ. Let $X_1, X_2,..., X_n$ be a random sample from X. Then the likelihood function is given as

$$L(\theta \mid X_1,..., X_n) = \prod_i f(X_i;\theta)$$

MLE can be obtained by differentiating the likelihood function with respect to the unknown parameters and putting the derivatives equal to 0. $\log L$ is a monotonically increasing function of L, so the value of the parameter that maximizes L also maximizes $\log L$.

MLE estimation is usually used econometrics for estimating Binary Choice and Limited Dependent Variable Models.

MLE is usually consistent and often unbiased.

MLE is generally the most asymptotically efficient estimator when the population model $f(y;\theta)$ is correctly specified.

MLE is sometimes the minimum variance unbiased estimator.

Let us apply this ML principle to the case of logit model for the regression $Y_i = \beta_1 + \beta_2 X_i + u_i$ where $Y_i = 1$ if the event takes place, $Y_i = 0$ otherwise and $E(u_i) = 0$.

By definition, likelihood function is

$$L = \prod_i p(Y = Y_i \mid X_i, \beta_1, \beta_2) = \prod_{i:Y_i=1} F(\beta_1 + \beta_2 \cdot X_i) \cdot \prod_{i:Y_i=0} (1 - F(\beta_1 + \beta_2 \cdot X_i)) \to \max_{\beta}$$

Instead of maximization of likelihood function usually use loglikelihood function

$$\log L(\beta_1, \beta_2) = \Sigma_i (\log p(Y = Y_i \mid X_i, \beta_1, \beta)) =$$

$$= \Sigma_{i:Y_i=1} \log F(\beta_1 + \beta_2 \cdot X_i) + \Sigma_{i:Y_i=0} \log(1 - F(\beta_1 + \beta_2 \cdot X_i))$$

The conditional summation in the last expression can be changed into unconditional one using coefficients Y_i and $1 - Y_i$ (first one takes value 1 for all elements of the first sum while the second takes value 1 for all elements of the second sum). So

$$\log L(\beta_1, \beta_2) = \sum_i [Y_i (\log F(\beta_1 + \beta_2 \cdot X_i)) + (1 - Y_i)(\log(1 - F(\beta_1 + \beta_2 \cdot X_i)))]$$

For logit model $\quad F(\beta_1 + \beta_2 \cdot X_i) = \dfrac{1}{1 + e^{-(\beta_1 + \beta_2 \cdot X_i)}}$

So $\log L(\beta_1, \beta_2) = \sum_i [Y_i (\log \dfrac{1}{1 + e^{-(\beta_1 + \beta_2 \cdot X_i)}}) + (1 - Y_i)(\log(1 - \dfrac{1}{1 + e^{-(\beta_1 + \beta_2 \cdot X_i)}}))]$

SECTION B

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

Question 4

4. [30 marks]. The relationship between a dependent variable Y and an explanatory variable X is given by the linear model

$$Y_i = \alpha + \beta X_i + u_i;\ i = 1, 2, \ldots, n$$

for an i.i.d. sample $\{Y_i, X_i\}$, but where the errors u_i are correlated with the explanatory variable, $\mathrm{E}(X_i u_i) = \sigma_{X,u} \neq 0$ (but $\mathrm{E}(X_i u_j) = 0$ for $i \neq j$).

(a) [10 marks]. Assuming that $\mathrm{plim}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right) = \sigma_X^2$, where $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, show that the OLS estimator of β is inconsistent.

Solution:

$$\hat{\beta}_{OLS} = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)} = \beta + \frac{\mathrm{Cov}(X,u)}{\mathrm{Var}(X)}$$

Taking plims we get

$$\mathrm{plim}\,\hat{\beta}_{OLS} = \beta + \frac{\mathrm{plim}\mathrm{Cov}(X,u)}{\mathrm{plim}\mathrm{Var}(X)} = \beta + \frac{\mathrm{cov}(X,u)}{\mathrm{var}(X)} = \beta + \frac{\sigma_{X,u}}{\sigma_X^2} \neq \beta, \text{ as } \sigma_{X,u} \neq 0.$$

Alternative solution: We have:

$$\mathrm{plim}(\hat{\beta}_{OLS} - \beta) = \mathrm{plim}\left(\frac{\sum(X_i - \overline{X})u_i}{\sum(X_i - \overline{X})^2}\right) = \left(\frac{\sum(X_i u_i - \overline{X}u_i)}{\sum(X_i - \overline{X})^2}\right) = \frac{\mathrm{plim}\left(\frac{1}{n}\left(\sum_i X_i u_i - \frac{1}{n}\sum_i\sum_j X_j u_i\right)\right)}{\mathrm{plim}\left(\frac{1}{n}\sum(X_i - \overline{X})^2\right)} = .$$

$$= \frac{\mathrm{plim}\left(\frac{1}{n}\left(\sum_i X_i u_i - \frac{1}{n}\sum_i X_i u_i - \frac{1}{n}\sum_{i\neq j} X_j u_i\right)\right)}{\mathrm{plim}\left(\frac{1}{n}\sum(X_i - \overline{X})^2\right)} = \frac{\mathrm{plim}\frac{1}{n}\sum_i X_i u_i - \left(\mathrm{plim}\frac{1}{n}\frac{1}{n}\sum_i X_i u_i + \mathrm{plim}\frac{1}{n}\frac{1}{n}\sum_{i\neq j} X_j u_i\right)}{\mathrm{plim}\left(\frac{1}{n}\sum(X_i - \overline{X})^2\right)} =$$

$$= \frac{\left(\mathrm{plim}\left(1 - \frac{1}{n}\right)\left(\frac{1}{n}\sum_i X_i u_i\right) - \mathrm{plim}\frac{1}{n}\frac{1}{n}\sum_{i\neq j} X_j u_i\right)}{\mathrm{plim}\left(\frac{1}{n}\sum(X_i - \overline{X})^2\right)} = \frac{\left(\mathrm{plim}\left(1 - \frac{1}{n}\right)\mathrm{plim}\left(\frac{1}{n}\sum_i X_i u_i\right) - \mathrm{plim}\frac{1}{n}\,\mathrm{plim}\frac{1}{n}\sum_{i\neq j} X_j u_i\right)}{\mathrm{plim}\left(\frac{1}{n}\sum(X_i - \overline{X})^2\right)} =$$

$= \frac{1\cdot \mathrm{E}(X_i u_i) - 0\cdot \mathrm{E}(X_j u_i)}{\sigma_X^2} = \frac{\sigma_{X,u}}{\sigma_X^2} \neq 0$ so the condition ($\mathrm{E}(X_i u_j) = 0$ for $i \neq j$) is excessive as it is not used in the proof.

(b) [5 marks]. Suppose now that only one additional exogenous variable Z is available in the data set. It is known that Z_i correlates with X_i, and does not already appear in the equation in its own right. Assume also that $\mathrm{E}(Z_i u_i) \neq 0$. Show that Z can not be considered as an instrument for X, as estimator $\hat{\beta} = \frac{\mathrm{Cov}(Y_i; Z_i)}{\mathrm{Cov}(X_i; Z_i)}$ of β is inconsistent.

Solution:
$\hat{\beta} = \frac{\text{Cov}(Y_i;Z_i)}{\text{Cov}(X_i;Z_i)} = \frac{\beta\,\text{Cov}(X_i;Z_i)+\text{Cov}(u_i;Z_i)}{\text{Cov}(X_i;Z_i)} = \beta + \frac{\text{Cov}(u_i;Z_i)}{\text{Cov}(X_i;Z_i)}$. Now taking plim we get

$\text{plim}(\hat{\beta}-\beta) = \frac{\text{plimCov}(u_i;Z_i)}{\text{plimCov}(X_i;Z_i)} = \frac{\text{cov}(u_i;Z_i)}{\sigma_{XZ}} = \frac{E(u_iZ_i)+E(u_i)E(Z_i)}{\sigma_{XZ}} = \frac{E(u_iZ_i)+0\cdot E(Z_i)}{\sigma_{XZ}} = \frac{E(u_iZ_i)}{\sigma_{XZ}} \neq 0$.

(c) [8 marks]. Three possible exogenous variables z_1, z_2 and z_3 are available in the data set. What properties do these potential instruments need to have to use them to estimate β consistently by the Instrumental Variable, or the Two Stage Least Squares (2SLS), estimators? Using all three instruments, describe in detail the 2SLS estimator

Solution:
Candidates need to explain that z_i must be correlated with X_i but not with u_i, and does not already appear in the equation in its own right. Hence in the reduced form:

$$X_i = \pi_0 + \pi_1 z_{i1} + \pi_2 z_{i2} + \pi_3 z_{i3} + \omega_i .$$

π_1, π_2 and π_3 should all not be zero (if one is zero then it should be dropped from the list). Furthermore, $\text{E}(z_i u_i) = 0$.

Estimate by OLS:

$$\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 z_{i1} + \hat{\pi}_2 z_{i2} + \hat{\pi}_3 z_{i3} .$$

Next, estimate by OLS:

$$Y_i = \alpha + \beta \hat{X}_i + \varepsilon_i$$

in the second stage. $\hat{X}$ only utilises exogenous variation due to the instruments, and hence identifies the causal effect β.

(d) [7 marks]. Explain in detail how would you test whether $\text{E}(X_i u_i) = 0$ (Durbin-Wu-Hausman test). Which estimator of β would you use if this hypothesis is not rejected?

Solution:
This is Durbin-Wu-Hausman test in the form of Davidson&MacKinnon
Add the first-stage residuals to the model:

$$Y_i = \alpha + \beta X_i + \gamma \hat{\omega}_i + \xi_i .$$

Testing for exogeneity is a test of $H_0 : \gamma = 0$. A t test is appropriate as standard errors are correct under the null hypothesis. If the test does not reject the null hypothesis use the OLS estimator of β.

Question 5

5. A student tries to find how the expenditure on education E_i (in billions of dollars) relates to Y_i - GDP (in billions of dollars) having data on 21 developed (such as Germany, France, Canada and so on) and developing countries (such as Uruguay, Chile, Mexico and so on) for the year 2010. First she runs simple regression model

$$\underset{(0.433)}{E_i = -0.589} + \underset{(0.007)}{0.069}Y_i + e_i \quad , \; R^2 = 0.84\,,\; i = 1, ..., 21, \qquad \textbf{(1)}$$

(a) [5 marks]. Give the interpretation to the slope coefficient. Why the student may fear of the presence of heteroscedasticity – explain in details on the basis of your understanding of heteroscedasticity. How heteroscedasticity could influence the regression results? How it can be seen on the graphs (what graphs)?

Solution:
The coefficient of GDP shows marginal effect of GDP on education expenditures. Each increase of GDP by 1 billion of dollars accompanies by increase of education expenditures by 68.6 millions of dollars. It is significant. In cross section samples of this type that includes developed and developing countries significantly different in their GDP, so it is naturally to expect the presence of heteroscedasticity: group of big countries usually have greater deviations (residuals), even if variations of education expenditures in different countries are the same in relative terms, they would be significantly different in absolute value.
Under heteroscedasticity the estimates being unbiased become inefficient, significance tests are invalid.
The heteroscedasticity can be visually observed on the scatter diagram as different range of observation scattering for different values of explanatory variable. It can be observed also on residual graph after sorting observation by explanatory variable: deviations of residuals from zero are generally greater at the certain parts of the graph (for example residuals become greater with the increase of explanatory variable).

(b) [10 marks]. Being worried of the presence of heteroscedastisity the student uses additionally option of White heteroscedasticity-consistent standard errors getting the following result: corrected standard error for the slope is now 0.01. Compare this result with the corresponding value in equation (1) and comment. Present and explain formula for heteroscedasticity-consistent standard errors for the slope coefficient and compare it with the corresponding formula under homoscedasticity for the standard errors for the slope coefficient.

Solution:
Under heteroscedasticity the standard errors in many cases are underestimated. Using the option of White heteroscedasticity-consistent standard errors allows to estimate them correctly. Obtained figures illustrate this: standard errors for the slop in equation (1) was 0.007 while corrected value 0.01 is greater, so we may conclude that significance of the slope in eq.(1) was slightly overestimated. It can be considered as one of the typical consequences of heteroscedasticity: standard errors of estimators increases, but sometimes it cannot be seen due to incorrect method of their estimation. White correction allows to fix this flaw.

Compare to methods of variance estimation: standard method with White correction.
Let us evaluate the variance of slope coefficient.

$$\sigma_{b_2}^2 = E\left\{(b_2 - E(b_2))^2\right\} = E\left\{(b_2 - \beta_2)^2\right\} = E\left\{\left(\sum_{i=1}^{n} a_i u_i\right)^2\right\} =$$

$$= E\left\{\sum_{i=1}^{n} a_i^2 u_i^2 + \sum_{i=1}^{n}\sum_{j\neq i} a_i a_j u_i u_j\right\} = \sum_{i=1}^{n} a_i^2 E(u_i^2) + \sum_{i=1}^{n}\sum_{j\neq i} a_i a_j E(u_i u_j) =$$

$$= \sum_{i=1}^{n} a_i^2 \sigma_i^2 .$$

Under homoscedasticity the variance is constant so $\sigma_i^2 = \sigma_u^2$, and so

$$\sigma_{b_2}^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2 = \sum_{i=1}^{n} a_i^2 \sigma_u^2 = \sigma_u^2 \sum_{i=1}^{n} a_i^2 = \frac{\sigma_u^2}{\sum_{j=1}^{n} (X_j - \overline{X})^2}$$

Under heteroscedastisity the expression will be different

$$\sigma_{b_2}^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2 == \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2 \sigma_i^2}{\left(\sum_{j=1}^{n} (X_j - \overline{X})^2\right)^2} = \frac{\sum_{i=1}^{n} x_i^{\,2} \sigma_i^2}{\left(\sum_{j=1}^{n} x_j^{\,2}\right)^2}, \text{ where } x_i = (X_i - \overline{X}).$$

Hence

$$s_{\hat{\beta}_2}^{\ \,2} = \frac{\sum_{i=1}^{n} x_i^{\,2} \hat{u}_i^2}{\left(\sum_{j=1}^{n} x_j^{\,2}\right)^2} = \sum_{i=1}^{n} a_i^{\,2} \hat{u}_i^2$$

Standard errors are biased but White (1980) shows that the following estimator is consistent: s.e.($b2$) =

$$s.e._{\hat{\beta}_2} = \sqrt{\frac{\sum_{i=1}^{n} x_i^{\,2} \hat{u}_i^2}{\left(\sum_{j=1}^{n} x_j^{\,2}\right)^2}} = \sqrt{\sum_{i=1}^{n} a_i^{\,2} \hat{u}_i^2} \quad \text{(heteroscedasticity-consistent standard error)}$$

Therefore, if it is impossible to identify the nature of heteroscedasticity, then in large samples heteroscedasticity-consistent standard errors make t-test and F-test asymptotically valid. However, there are some problems:

1) The obtained estimator may not perform well in finite samples;

2) OLS point estimates remain inefficient.

(c) [7 marks]. She decided to add as explanatory variable also P_i - population of each country (in millions). The regression now is

$$\underset{(0.402)}{E_i = -0.224} + \underset{(0.006)}{0.073\, Y_i} - \underset{(0.025)}{0.073\, P_i} + e_i, \quad R^2 = 0.89, \ i = 1, ..., 21, \qquad \textbf{(2)}$$

with the sample covariance between Y_i and P_i being positive.

The student was stumped that the coefficient of Y_i has changed. Try to explain her that both the direction and the absolute value of this change are in strict accordance with the theory of specification of a regression model. How can be explained negative sign of the coefficient of P_i?

Included additionally variable P_i is significant ($\frac{-0.073}{0.025} = 2.92$ while $t(crit.,1\%.,df = 18) = 2.878$) that can be considered as a strong argument in favor of specification of equation (2). Under assumption that specification

$$E_i = \beta_1 + \beta_2 Y_i + \beta_3\ P_i + u_i$$

corresponding to equation **(2)** is true, the slope coefficient of equation **(1)**

$$\hat{E}_i = \hat{\beta}_1 + \hat{\beta}_2 Y_i$$

is biased

$$\hat{\beta}_2 = \beta_2 + \beta_3 \frac{\text{Cov}(Y_i, P_i)}{\text{Var}(Y_i)}$$

As specification **(2)** is supposed to be correct and the estimate of coefficient β_3 is significant and negative ($\hat{\beta}_3 = -0.073$) the true value of coefficient β_3 is likely also negative. As we are told that $\text{Cov}(Y_i, P_i) > 0$, the sign of the bias $\beta_3 \dfrac{\mathrm{Cov}(Y_i, P_i)}{\mathrm{Var}(Y_i)}$ is negative. So the coefficient $\hat{\beta}_2$ in specification (1) is biased downwards. In fact $0.069 < 0.073$.

The reason for coefficient $\hat{\beta}_3 = -0.073$ to be negative is that many most populated countries are among the developing countries so they can afford themselves smaller expenditures on education.

(d) [8 marks]. The student decided to test her regression (2) for heteroscedastisity. First she attemted to perform Goldfeld-Quandt test: she ordered all observations by the values of the variable Y_i (in ascending order) and then run the regression in specification (2) using two different samples: 1st – 7 developing countries with lowest values of GDP, and 2nd – 10 developed countries with greatest values of GDP, getting two ifferent values of RSS – 0.490195 for the 1st sample, and 15.24647 for the 2nd sample. She also runs two different White tests with the following results

1 Heteroskedasticity White Test (do not include White cross term):
Equation: Dependent Variable: RESID^2, Expanatory variables: C, Y^2, P^2, R-squared=0.42302

2) Heteroskedasticity White Test(include White cross term):
Equation: Dependent Variable: RESID^2, Expanatory variables: C, Y, Y^2, Y*P, P, P^2; R-squared=0.483819.

Help the student to continue performing all three tests (Goldfeld-Quandt and two version of White test), comment them and explain the difference between their results and make a final conclusion. If according to your conclusion the heteroscedastisity is a problem here suggest some solutions to it.

To finalize Goldfeld-Quandt test it is sufficient to evaluate F-statistics $F = \dfrac{15.24647/7}{0.490195/4} = 17.773$ and compare it with critical value of $F(crit.,1\%., df1 = 7, df1 = 4) = 14.98$. The null hypothesis of homoscedasticity is rejected at 1% significance level.
The situation with White test is more complicated. Test statistic here is $n \cdot R^2$ that has chi-square distribution with the degrees of freedom equal to the number of variables in the auxillary equation for squared residuals.
So for White test that does not include cross terms test statustics is $21 \cdot 0.423020 = 8.8834$ with the critical value of $\chi^2(1\%, df = 2) = 9.21$ and $\chi^2(5\%, df = 2) = 5.99$ so the null hypothesis of homoscedasticity is rejected only at 5% significance level.
For White test with inclusion of cross terms test statustics is $21 \cdot 0.483819 = 10.1602$ with the critical value of $\chi^2(1\%, df = 5) = 15.0863$ and $\chi^2(5\%, df = 5) = 11.0705$ so the null hypothesis of homoscedasticity is not rejected.

Discussion and conclusion
Goldfeld-Quandt test is specialized on testing for heteroscedasticiy of the certain type, when the standard deviation of the disturbance term of the regression is proportional to some explanatory variable. So the obtained result certainly indicates on the presence of heteroscedasticity of this type: in this case disturbance term of the regression is proportional to GDP.

White test is more flexible and universal: it allows to detect heteroscedasticity of different type including those caused by combined action of two and more variables. But it does not mean that it is always better that specialized test. In our case White test without cross terms is significant only at 5% significance level. As it is indicated in the question this test uses only Y^2 and P^2 terms what does not correspond properly to the heteroscedasticity caused by propotionality factor Y (GDP) detected by Goldfeld-Quandt test.
White test with cross terms works even worse: too mane factors included for just 21 observations, possible multicollinearity prevent detection of heteroscedastisity.
As factor proportionality is found WLS should be recommended using GDP as a weihting series.

The International College of Economics and Finance
Econometrics - 2017. First Semester Exam, December 28.
Suggested Solutions

Part 2. (1 hour 30 minutes). Answer all questions (1,2,3) from section A and one (4 or 5) - from section B.

IMPORTANT: Start answering each question from the new page (ask for extra paper if necessary). Structure your answers in accordance with the structure of the questions. Testing hypotheses always state clearly null and alternative hypotheses provide critical value used for test, mentioning degrees of freedom and the significance level chosen for the test.

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

1. [15 marks]. A student decided to investigate the market of private mathematics teachers in Moscow, with particular interest to those who can teach in English. He took a random sample of 30 profiles of teachers who provide private teaching in math (taken from population of 300 profiles registered in certain internet site) and run some regressions trying to find factors influencing the prices of teaching ($PRICE_i$ - price of a standard two-hour lesson in thousands of roubles, $DIST_i$ - distance in the number of metro stations from the center of Moscow to the teacher's place, $HOME_i$ - dummy variable indicating visit of the tutor to the client, ENG_i - dummy variable indicating ability to teach the subject in English):

$$\hat{PRICE_i} = \underset{(0.49)}{6.59} - \underset{(0.06)}{0.16}DIST_i \qquad R^2 = 0.185 \tag{1}$$

$$\hat{PRICE_i} = \underset{(0.40)}{4.51} + \underset{(0.58)}{2.54}HOME_i \qquad R^2 = 0.40 \tag{2}$$

$$\hat{PRICE_i} = \underset{(0.64)}{5.13} - \underset{(0.06)}{0.08}DIST_i + \underset{(0.95)}{1.95}HOME_i + \underset{(0.07)}{0.07}DIST*HOME_i \qquad R^2 = 0.437 \tag{3}$$

$$\hat{PRICE_i} = \underset{(0.61)}{4.52} - \underset{(0.06)}{0.08}DIST_i + \underset{(0.75)}{2.18}HOME_i + \underset{(0.76)}{1.58}ENG_i - \underset{(1.09)}{0.39}HOME*ENG_i \qquad R^2 = 0.553 \tag{4}$$

(a) [5 marks]. What is the meaning of equations (1) and (2)? What are advantages and disadvantages of these equations? What is the difference in the meaning and in the assumptions (explicit or implicit) used for equation (3) as to compare to equation (1)? Give interpretation to all coefficients of equation (4).

Equation (1). The coefficient $-0,16$ shows that with each additional metro station in distance from the center of Moscow the price of the service drops on average by 160 roubles, while if the teacher lives in the center the average price is 6.59 thousand roubles. Both values do not take into account whether the service is provided in teacher's place or at the place of the student. So the coefficient of this equation could suffer of the omitted variable bias. For this reason all tests for this equation can be invalid.
Equation (2). The constant term 4.51 shows average price of the teaching at the teacher's place in apartments, while the sum of two coefficients 4.51+2.54=7.1 shows the average price in thousand of roubles for the teaching at the student's place. This is so called degenerate equation, as it contains only dummy variable and no control variable. It adds no information to the simple statistical characteristics of the data.

Equation (3). The constant term 5.13 now shows the average price of the teaching at the teacher's place, in this equaion the influence of the distance is different for the teaching at the teacher's place and at home: the equation for the teaching at the teacher's place is

$$\hat{PRICE_i} = 5.13 - 0,08DIST_i,$$

while for service out of department the equations is

$$\hat{PRICE_i} = (5.13 + 1,95) + (-0,08_i + 0,07)DIST_i$$

or

$$\hat{PRICE_i} = 7.08 - 0,01DIST_i.$$

Equation (4). The constant term now is the price of the teaching of math teacher that gives lessons at his home, living in the center and not able to teach math in English ($DIST = 0, HOME = 0, ENG = 0$). Each additional metro station in distance from center lowers the price by 80 roubles not taking into account the possibility of teaching at the student's place or ability to teach in English. Teaching at the student's place keeping other factors constant, adds to the price 2.18 thousand of roubles. Additional ability to teach in English adds $1.58 \cdot 1 - 0.39 \cdot 1 \cdot 1 = 1.19$ thousand roubles to the price of teaching.

(b) [5 marks]. Is equation (3) significant? Is factor "distance" (the variables $DIST_i$ and $DIST * HOME_i$ taken together) significant in equation (3)? Is factor "teaching at student's place" (the variables $HOME_i$ and $DIST * HOME_i$ taken together) significant in equation (3)? Are all dummy variables taken together in equation (4) significant?

To estimate significance of the equation (3) one can use F-test:

$$F = \frac{0.437/3}{(1-0.437)/(30-4)} = 6.73$$

with critical value of $F_{crit}(3, 26, 1\%) = 4.64$, so the hypothesis of $H_0: \beta_1 = 0, \beta_2 = 0, \beta_3 = 0$ for equation

$$PRICE = \beta_0 + \beta_1 DIST + \beta_2 HOME + \beta_3 DIST * HOME + u_t \quad (3^*)$$

is rejected at 1% significance level.

The significance of $DIST_i$ and $HOME_i$ does not follow from here. To test it we should use another tests based on the comparison of equation (3) with (1) and (2).

The equation

$$PRICE = \beta_0 + \beta_1 DIST + u_t \quad (1^*)$$

is a restricted version of equation (3*) with the restriction $H_0: \beta_2 = 0, \beta_3 = 0$, so $F = \frac{(0.437-0.185)/2}{(1-0.437)/(30-4)} = 5.82$ with critical value of $F_{crit}(2, 26, 1\%) = 5.53$, so the null hypothesis is rejected at 1%.

The equation

$$PRICE = \beta_0 + \beta_2 HOME + v_t \quad (2^*)$$

is a restricted version of equation (3*) with the restriction $H_0: \beta_1 = 0, \beta_3 = 0$, so $F = \frac{(0.437-0.40)/2}{(1-0.437)/(30-4)} = 0.85$ with critical value of $F_{crit}(2, 26, 5\%) = 3.37$, so the null hypothesis is not rejected.

The equation $PRICE = \beta_0 + \beta_1 DIST + u_t$ (1*) is a restricted version of equation (4*)

$$\widehat{PRICE}_i = \beta_0 + \beta_1 DIST_i + \beta_2 HOME_i + \beta_3 ENG_i + \beta_4 HOME \cdot ENG_i + w_i$$

with the restriction $H_0: \beta_2 = 0, \beta_3 = 0, \beta_4 = 0$, so $F = \frac{(0.553-0.185)/3}{(1-0.553)/(30-5)} = 6.86$ with critical value of $F_{crit}(3, 25, 1\%) = 4.68$, so the null hypothesis is rejected at 1%.

(c) [5 marks]. Another approach to the analysis of influence of qualitative factors is Chow test. Explain how you will do Chow test for the analysis of the influence of the place of teaching. Can you predict the result of this test? Explain how you will do Chow test for the analysis of the influence of both the place of teaching and ability to teach in English. Write the equation with dummy variables $H = HOME$ and $E = ENG$ that allows to do F-test for restrictions that is equivalent to the last of the two Chow tests.

To do Chow test that is equivalent to the F-test for the hypothesis $H_0: \beta_2 = 0, \beta_3 = 0$ in (b), one should estimate equation (1*)

$$PRICE = \beta_0 + \beta_1 DIST + u_t \quad (1^*)$$

for the whole sample memorising *RSS*(*total*), and then repeat this for two sub-samples for teacher's place teaching and student's place teaching getting correspondingly *RSS*(*t*) and *RSS*(*s*) and then evaluate F-statistic

$$F = \frac{(RSS(total) - (RSS(t) + RSS(s)) / 2}{(RSS(t) + RSS(s) / (30 - 2 \cdot 2)}$$

It will be equal to 5.82 with the same characteristics of significance.

To take into account two different factors (place of teaching and Enlish) we need to specify 4 take subsamples, consisting of teachers that 1) teach at teacher's place in Russian – (tR), 2) teach at teacher's place in Russia – (tR), 3) teach at student's place in Russian – (sR), and 4) teach at student's place in English – (tE); estimate equation

$$PRICE = \beta_0 + \beta_1 DIST + u_t \ \ (1^*)$$

for the whole sample getting *RSS*(*total*), and then repeat this for 4 sub-samples obtaining *RSS*(*tR*), *RSS*(*tE*), *RSS*(*sR*) and *RSS*(*sE*), and then evaluate F-statistic

$$F = \frac{(RSS(total) - (RSS(tR) + RSS(tE) + RSS(sR) + RSS(sE)) / (2 \cdot 3)}{(RSS(tR) + RSS(tE) + RSS(sR) + RSS(sE)) / (30 - 2 \cdot 4)}$$

Alternative way is to estimate equation (denoting $HOME$ as H and ENG as E

$$\hat{PRICE}_i = \beta_0 + \beta_1 DIST_i + \beta_2 H_i + \beta_3 E_i + \beta_4 H \cdot E_i + \beta_1 DIST \cdot H_i + \beta_3 DIST \cdot E_i + \beta_4 DIST \cdot H \cdot E_i + w_i \ \ (5^*)$$

memorise its $R^2(5^*)$ and then evaluate F-test

$$F = \frac{(R^2(1^*) - R^2(5^*) / 6}{(1 - R^2(5^*)) / (30 - 8)}$$

that is equivalent to the last Chow test.

2. [15 marks]. Consider simple linear regression model without constant

$$y_t = \alpha x_t + u_t; \ t = 1, 2, ..., T$$

x_t is an observed non-random variable. We assume that $\mathrm{E}(u_t) = 0$; $\mathrm{E}(u_s u_t) = 0$ if $s \neq t$, for all *s* and *t*, and normally distributed, but the values of disturbance term u_t are characterized by heteroscedastisity of a certain type $\mathrm{E}(u_t^2) = \sigma^2 x_t^2$, so the density function of u_t is different for each observation is $f(u_t) = \frac{1}{\sqrt{2\pi\sigma^2 x_t^2}} \cdot e^{-\frac{1}{2}\left(\frac{y_t - \alpha x_t}{\sigma x_t}\right)^2}$.

(a) [5 marks] What is the maximum likelihood principle in relation to the problem under consideration.? Derive the log likelihood function for the model under consideration to find parameters α and σ^2 given the sample values $L(\alpha, \sigma^2 | u_1, u_2.,..u_T)$.

According maximum likelihood principle parameters of regression α, σ^2 should be chosen in such a way as to maximize the probability of obtaining the sample values of disturbance term $u_t, t = 1, 2, ..., T$.

To derive ML estimator we should first construct the likelihood function

$$L(\alpha, \sigma^2 | u_1, u_2.,..u_n) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2 x_t^2}} \cdot e^{-\frac{1}{2}\left(\frac{y_t - \alpha x_t}{\sigma x_t}\right)^2}.$$

Log likelihood function becomes (all logarithms are assumed to be natural):

$$\log L(\alpha, \sigma^2 | u_1, u_2.,..u_n) = \sum_{t=1}^{T} \log\left(\frac{1}{\sqrt{2\pi\sigma^2 x_t^2}} \cdot e^{-\frac{1}{2}\left(\frac{y_t - \alpha x_t}{\sigma x_t}\right)^2}\right) = .$$

$$= -\frac{T}{2}\log 2\pi - \frac{1}{2}\log\sum_{t=1}^{T}\sigma^2 x_t^2 - \frac{1}{2}\sum_{t=1}^{T}\left(\frac{y_t - \alpha x_t}{\sigma x_t}\right)^2 = -\frac{T}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2}\sum_{t=1}^{T} x_t^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(\frac{y_t}{x_t} - \alpha\right)^2$$

From the obtained expression, maximization of the log likelihood function implies minimization of RSS for choosing estimator of β, Hence, ML estimator of β coincide with the OLS one.

(b) **[5 marks]** Using likelihood function from **(a)** derive maximum likelihood (*ML*) estimators of α and σ^2. Show that $\hat{\alpha}$ the *ML* estimator of α is unbiased.

Solution:
The first-order conditions are:

$$\frac{\partial \log L}{\partial \alpha} = -\frac{1}{2\sigma^2}\sum\left(\frac{y_t}{x_t}-\alpha\right)=0 \qquad \text{(i)}$$

$$\frac{\partial \log L}{\partial \sigma^2} = -\frac{T}{2\sigma^2}+\frac{1}{2\sigma^4}\sum\left(\frac{y_t}{x_t}-\alpha\right)^2=0. \qquad \text{(ii)}$$

Solving (i) and (ii), the *ML* estimators of α and σ^2 are obtained as:

$$\hat{\alpha}_{MLE}=\frac{1}{T}\sum\left(\frac{y_t}{x_t}\right)$$

Then $\frac{T}{2\sigma^2}=\frac{1}{2\sigma^4}\sum\left(\frac{y_t}{x_t}-\alpha\right)^2$, so $T=\frac{1}{\sigma^2}\sum\left(\frac{y_t}{x_t}-\alpha\right)^2$ and so

$$\hat{\sigma}^2_{MLE}=\frac{1}{T}\sum\left(\frac{y_t}{x_t}-\alpha\right)^2.$$

To show $\hat{\alpha}_{MLE}$ is an unbiased estimator,

$$\mathrm{E}(\hat{\alpha}_{MLE})=\frac{1}{T}\sum\left(\frac{\mathrm{E}(y_t)}{x_t}\right)=\frac{1}{T}\sum\left(\frac{\alpha x_t}{x_t}\right)=\frac{\alpha}{T}\sum\left(\frac{x_t}{x_t}\right)=\frac{T\alpha}{T}=\alpha$$

Hence $\hat{\alpha}_{MLE}$ is an unbiased estimator of α.

(c) **[5 marks]** Compare the *ML* estimator of α, with the weighted least squares estimator of α.

To obtain the weighted least squares (WLS) estimator, first the model is divided by x_t to get $\frac{y_t}{x_t}=\alpha+\frac{u_t}{x_t}$.

Denote $y_t^*=\frac{y_t}{x_t}=\alpha+\frac{u_t}{x_t}$, $x_t^*=1$, $u_t^*=\frac{u_t}{x_t}$, so $y_t^*=\alpha x_t^*+u_t^*$, OLS estimator for regression without constant:

$$\hat{\alpha}=\frac{\sum y_t^* x_t^*}{\sum (x_t^*)^2}=\frac{\sum \frac{y_t}{x_t}\cdot 1}{\sum (1)^2}=\frac{1}{T}\sum\frac{y_t}{x_t}.$$

Alternatively to get WLS estimator for α one should find such value of $\hat{\alpha}_{WLS}$ that sum $\sum e^2$ is minimized for the model $\frac{y_t}{x_t}=\hat{\alpha}_{WLS}+e_t$. So $\sum e^2=\sum\left(\frac{y_t}{x_t}-\hat{\alpha}_{WLS}\right)^2$ has to be minimized.

F.O.C. are $-2\sum\left(\frac{y_t}{x_t}-\hat{\alpha}_{WLS}\right)=0$ and so $\hat{\alpha}_{WLS}=\frac{1}{T}\sum\left(\frac{y_t}{x_t}\right)$

Hence WLS estimators and *ML* estimators are the same.

3. [**15 marks**] The researcher is trying to apply short-term equilibrium model (Mundell-Flemming model) for exploring data on small European country with open economy.

$$Y_t = C_t + I_t + NX_t \qquad (1) \qquad \text{- } \textit{macroeconomic identity}$$
$$C_t = \alpha + \beta Y_t + u_t \qquad (2) \qquad \text{- } \textit{consumption function}$$
$$I_t = \delta - \varepsilon R_t + \lambda Y_t + v_t \qquad (3) \qquad \text{- } \textit{investment function}$$
$$NX_t = \gamma - \eta Y_t - \varphi E_t + w_t \qquad (4) \qquad \text{- } \textit{net exports function}$$
$$(M/P)_t = \mu Y_t - \theta R_t + s_t \qquad (5) \qquad \text{- } \textit{money market equation}$$

where income Y_t, consumption C_t, investment I_t, net exports NX_t and exchange rate E_t are the endogenous variables. Variables R_t (interest rate which is set at the world level) and $(M/P)_t$ (real money supply) are exogenous; u_t, v_t, w_t, s_t are the disturbance terms. Indicate the correct statement:

(a) [**5 marks**] Putting aside all economic problems connected with Mundell-Flemming model determine which of the equations (2)-(4) are exactly identified, which are over-identified, and which are underidentified.

There are two ways to judge on identification of each equation: to find a number of instruments for some endogenous variables, or to use order condition (this is only necessary condition but sufficient conditions are out of the scope of our course). To apply order condition we should know the total number of equations G (that is always equal to the number of endogenous variables, here $G = 5$. Total number of variables (both endogenous and exogenous) is 7. If the number of variables missing from equation is equal to $G-1$, the equation is likely to be exactly identified. If it is less than $G-1$ the equation is likely to be underidentified. If it is more than $G-1$ the equation is likely to be overidentified. In our case $G-1=4$.
Let us apply this rule to equations under consideration in turn.

$$Y_t = C_t + I_t + NX_t \qquad \text{- } \textit{macroeconomic identity}$$

Identification is not applicable.

$$C_t = \alpha + \beta Y_t + u_t \qquad \text{- } \textit{consumption function}$$

5 variables missed from equation => overidentified.
(Alternatively: there are two possible instruments R_t and $(M/P)_t$ for the endogenous variable Y_t to be instrumented).

$$I_t = \delta - \varepsilon R_t + \lambda Y_t + v_t \qquad \text{- } \textit{investment function}$$

4 variables missed from equation => exactly identified.
(Alternatively: there is one possible instrument $(M/P)_t$ for endogenous variable Y_t to be instrumented).
The same situation in equation

$$(M/P)_t = \mu Y_t - \theta R_t + s_t \qquad \text{- } \textit{money market equation}$$

The situation here is rather special and will be discussed in **c)**.

$$NX_t = \gamma - \eta Y_t - \varphi E_t + w_t \qquad \text{- } \textit{net exports function}$$

4 variables missed from equation => exactly identified.
(Alternatively: there are two possible instruments R_t and $(M/P)_t$ for two endogenous variables Y_t and R_t to be instrumented).

(b) [**5 marks**] What method is most appropriate for estimation of investment function? What method is most appropriate for estimation of consumption function? Describe supposed methods in details.

As investment function is exactly identified. So any method giving consistent estimates will be appropriate. For example the only available instrument $(M/P)_t$ can be used to obtain IV estimation for the parameter λ. It is also possible to use TSLS or even to try to get ILS estimator (but it could be a little tedious work).
As consumption function is overidentified two possible instruments are available, both are consistent but different, so TSLS is the best solution to find a compromise for them.
At the first stage we use both instrument to estimate equation using OLS

$$Y_t = \rho_1 + \rho_2 R_t + \rho_3 (M/P)_t + \sigma_t$$

and then use obtained estimates $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3$ to evaluate the most efficient (as it is obtainted by OLS) instrument

$$\hat{Y}_t = \hat{\rho}_1 + \hat{\rho}_2 R_t + \hat{\rho}_3 (M/P)_t$$

This 'superinstrument' then is used at the second stage to get consistent estimates of the parameters of equation

$$C_t = \alpha + \beta \hat{Y}_t + u_t$$

also using OLS.

(c) [**5 marks**] How can the researcher estimate equation (5)? What properties are expected to have the obtained estimates?

The situation here is special. According order condition money market equation (5) is exactly identified. But the presence of the exogenous variable $(M/P)_t$ in the left side of the equation is a serious problem. To estimate this equation we have to swap the variables $(M/P)_t$ and Y_t to have endogenous variable at the left side of the estimated equation

$$Y_t = \psi_1 (M/P)_t + \psi_2 R_t + \omega_t$$

As right-hand side of this equation contains only exogenous variables $(M/P)_t$ and R_t OLS gives consistent estimates of parameters $\hat{\psi}_1, \hat{\psi}_2$. But it is impossible to get from here consistent estimates for the original equation. To explain this for the sake of clarity consider two simple linear regressions

$$Y_t = \alpha X_t + u_t \quad (*)$$

and

$$X_t = \beta Y_t + v_t \quad (**)$$

OLS estimators are $\hat{\alpha} = \dfrac{\sum X_t Y_t}{\sum X_t^2}$ and $\hat{\beta} = \dfrac{\sum X_t Y_t}{\sum Y_t^2}$. Generally speaking they cannot be dirived from each other.

To get $\hat{X}_t = bY_t$ from $\hat{Y}_t = aX_t$ we should take reciprocal of a: $b = \dfrac{1}{a}$. But the equality

$$\frac{1}{\hat{\alpha}} = \frac{1}{\dfrac{\sum X_t Y_t}{\sum X_t^2}} = \frac{\sum X_t^2}{\sum X_t Y_t} = \frac{\sum X_t Y_t}{\sum Y_t^2} = \hat{\beta}$$

means that $\dfrac{\sum X_t^2}{\sum X_t Y_t} = \dfrac{\sum X_t Y_t}{\sum Y_t^2}$ or $\dfrac{(\sum X_t Y_t)^2}{\sum X_t^2 \sum Y_t^2} = 1$ or $r_{x,y}^2 = 1$, that means that there is perfect linear dependence between X_t and Y_t.

It can be explained also by graphical illustration: in estimation of equation (*) OLS uses deviations from regression line along the Y-axis, while for estimation of equation (**) the deviations are evaluated along X-axis, so the results of the estimation are different. The equation (5) is multiple regression, so the situation is more complicated but the general result should be the same.

4. [30 marks] In year *t*, aggregate demand for a certain commodity, Q_{Dt}, is related to its price, P_t, and also to the aggregate income, Y_t, which is supposed to be exogenous:

$$Q_{Dt} = \beta_1 + \beta_2 P_t + \beta_3 Y_t + u_{Dt} \qquad (1)$$

Aggregate supply in year *t*, Q_{St}, is also a function of P_t:

$$Q_{St} = \alpha_1 + \alpha_2 P_t + u_{St} \qquad (2)$$

u_{Dt} and u_{St} are disturbance terms that satisfy the Gauss–Markov conditions and are distributed independently of each other. The market clears in each year, so that $Q_{Dt} = Q_{St}$.

For the purposes of this question, any problems associated with nonstationary time series may be ignored.

(a) [5 marks] Explain why OLS would not be appropriate for estimation parameters of both equations. (No estimation of large sample bias is expected and no marks will be given for this). What are exogenous and endogenous variables? What is the difference between econometric equations in the structural and reduced form? Derive reduced form equations.

Violation of GMC

Since P is partly determined by both u_{Dt} and u_{St}, the Gauss–Markov condition that the explanatory variables are distributed independently of the disturbance term is violated for both equations and as a consequence OLS yields inconsistent estimates.

Reduced form equations

To derive the reduced form equation for *P* it is sufficient to deduct equation (2) from equation (1) taking into account equality $Q_{Dt} = Q_{St}$

$$Q_{Dt} = \beta_1 + \beta_2 P_t + \beta_3 Y_t + u_{Dt} \qquad (1)$$

$$Q_{St} = \alpha_1 + \alpha_2 P_t + u_{St} \qquad (2)$$

$$Q_{Dt} - Q_{St} = (\beta_1 - \alpha_1) + (\beta_2 - \alpha_2)P_t + \beta_3 Y_t + u_{Dt} - u_{St},$$

so the reduced form equation for *P* is

$$P_t = \frac{\beta_1 - \alpha_1}{\alpha_2 - \beta_2} + \frac{\beta_3 Y_t}{\alpha_2 - \beta_2} + \frac{u_{Dt} - u_{St}}{\alpha_2 - \beta_2}$$

To get the reduced form equation for *Q* it is convenient to multiply equation (1) by α_2, equation (2) – by β_2 and then subtract them taking into account that $Q_{Dt} = Q_{St} = Q_t$

$$\alpha_2 Q_t = \alpha_2\beta_1 + \alpha_2\beta_2 P_t + \alpha_2\beta_3 Y_t + \alpha_2 u_{Dt} \qquad (1')$$

$$\beta_2 Q_t = \alpha_1\beta_2 + \alpha_2\beta_2 P_t + \beta_2 u_{St} \qquad (2')$$

$$(\alpha_2 - \beta_2)Q_t = (\alpha_2\beta_1 - \alpha_1\beta_2) + \alpha_2\beta_3 Y_t + (\alpha_2 u_{Dt} - \beta_2 u_{St}) \qquad (1') - (2')$$

So

$$Q_t = \frac{\alpha_2\beta_1 - \alpha_1\beta_2}{\alpha_2 - \beta_2} + \frac{\alpha_2\beta_3}{\alpha_2 - \beta_2} Y_t + \frac{\alpha_2 u_{Dt} - \beta_2 u_{St}}{\alpha_2 - \beta_2}$$

(b) [7 marks] What is ILS (indirect least squares) method? How it can be used to estimate parameters of the equation (2)?

The equations of reduced form system could be rewritten as

$$P_t = \pi_{11} + \pi_{12} Y_t + w_{1t} \qquad (4)$$

$$Q_t = \pi_{21} + \pi_{22} Y_t + w_{2t} \qquad (5)$$

Where $\pi_{11} = \dfrac{\beta_1 - \alpha_1}{\alpha_2 - \beta_2}$ (5); $\pi_{12} = \dfrac{\beta_3}{\alpha_2 - \beta_2}$ (6); $\pi_{21} = \dfrac{\alpha_2\beta_1 - \alpha_1\beta_2}{\alpha_2 - \beta_2}$ (7); $\pi_{22} = \dfrac{\alpha_2\beta_3}{\alpha_2 - \beta_2}$ (8);

And also $w_{1t} = \dfrac{u_{Dt} - u_{St}}{\alpha_2 - \beta_2}$; $w_{2t} = \dfrac{\alpha_2 u_{Dt} - \beta_2 u_{St}}{\alpha_2 - \beta_2}$.

The equations of the system (3)-(4) can be estimated using OLS (getting BLUE estimators $\hat{\pi}_{11}, \hat{\pi}_{12}, \hat{\pi}_{21}, \hat{\pi}_{22}$,) as each equation includes only exogenous variables in its right hand side.

It is easily could be seen that $\alpha_2 = \frac{\pi_{22}}{\pi_{12}}$ as $\frac{\pi_{22}}{\pi_{12}} = \frac{\frac{\alpha_2\beta_3}{\alpha_2 - \beta_2}}{\frac{\beta_3}{\alpha_2 - \beta_2}} = \alpha_2$

Also $\alpha_1 = \pi_{21} - \frac{\pi_{22}}{\pi_{12}}\pi_{11}$ as

$$\pi_{21} - \frac{\pi_{22}}{\pi_{12}}\pi_{11} = \frac{\alpha_2\beta_1 - \alpha_1\beta_2}{\alpha_2 - \beta_2} - \alpha_2\frac{\beta_1 - \alpha_1}{\alpha_2 - \beta_2} = \frac{\alpha_2\beta_1 - \alpha_1\beta_2 - \alpha_2\beta_1 + \alpha_2\alpha_1}{\alpha_2 - \beta_2} = \alpha_1\frac{\alpha_2 - \beta_2}{\alpha_2 - \beta_2} = \alpha_1$$

So it is possible to estimate all coefficients of equation (2): $\alpha_2^{ILS} = \frac{\hat{\pi}_{22}}{\hat{\pi}_{12}}$, $\alpha_1^{ILS} = \hat{\pi}_{21} - \frac{\hat{\pi}_{22}}{\hat{\pi}_{12}}\hat{\pi}_{11}$.

(c) [5 marks] Prove that ILS estimators are consistent. What are other methods providing consistent estimators for it?

As it has been said the estimators $\hat{\pi}_{11}, \hat{\pi}_{12}, \hat{\pi}_{21}, \hat{\pi}_{22}$ are BLUE. The same cannot be said on estimators $\alpha_2^{ILS} = \frac{\hat{\pi}_{22}}{\hat{\pi}_{12}}$, $\alpha_1^{ILS} = \hat{\pi}_{21} - \frac{\hat{\pi}_{22}}{\hat{\pi}_{12}}\hat{\pi}_{11}$ as their evaluation involve division by stochastic variables so it is impossible to use expectations to prove unbiasedness. Nevertheless it is possible to use plims to investigate their large sample properties.

These estimators are consistent as $\operatorname{plim}\alpha_2^{ILS} = \frac{\operatorname{plim}\hat{\pi}_{22}}{\operatorname{plim}\hat{\pi}_{12}} = \frac{\pi_{22}}{\pi_{12}} = \alpha_2$,

and $\operatorname{plim}\alpha_1^{ILS} = \operatorname{plim}\hat{\pi}_{21} - \frac{\operatorname{plim}\hat{\pi}_{22}}{\operatorname{plim}\hat{\pi}_{12}}\operatorname{plim}\hat{\pi}_{11} = \pi_{21} - \frac{\pi_{22}}{\pi_{12}}\pi_{11} = \alpha_1$.

Another appriaches to get consistent estimates are IV estimators and TSLS.

(d) [5 marks] What is meant by identification of equations? What ILS can tell on the identification of the original equations in structural form? What are alternative methods to tell whether an equation from the system of econometric equations is identified?

The equation is said to be identified if there is a method to get consistent estimates of its parameters. From the existence of a set of consistent estimators for equation (2) in (c) follows that the equation (2) is exactly identified. As it is impossible to find 5 parameters $\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3$ from 4 equations the equation (1) is underidentified.

Another approaches to identification is IV method and order condition. One can use variable Y_t as an instrument for endogenous variable P_t in equation (2), so equation (2) is (exactly) identified. The only exogenous variable Y_t already is used in equation (1) as an explanatory variable and so cannot be used there as an instrument, so the equation (1) is underidentified.

To use order condition for identification we shoud first to know the number of equations or alternatively the number of endogenous variables, it is $G = 2$. So the main parameter is $G-1=1$. As only one variable is missing from equation (2), this equation is exactly identified. None variables are missing from equation (1) so it is underidentified.

(e) [8 marks] Prove that all three methods ILS (Indirect Least Squares), IV (Instrumental Variables) and TSLS (Two Stage Least Squares) used for estimation of the α_2 of equation (2) give the identical results.

$$Q_{Dt} = \beta_1 + \beta_2 P_t + \beta_3 Y_t + u_{Dt} \qquad (1)$$
$$Q_{St} = \alpha_1 + \alpha_2 P_t + u_{St} \qquad (2)$$

or

$$P_t = \pi_{11} + \pi_{12} Y_t + w_{1t} \qquad (3)$$
$$Q_t = \pi_{21} + \pi_{22} Y_t + w_{2t} \qquad (4)$$

$$\hat{\alpha}_2^{ILS} = \frac{\pi_{22}}{\pi_{12}} = \frac{\dfrac{\mathrm{Cov}(Y,Q)}{\mathrm{Var}(Y)}}{\dfrac{\mathrm{Cov}(Y,P)}{\mathrm{Var}(Y)}} = \frac{\mathrm{Cov}(Y,Q)}{\mathrm{Cov}(Y,P)} = \hat{\alpha}_2^{IV}$$

$$\hat{\alpha}_2^{ILS} = \frac{\pi_{22}}{\pi_{12}} = \frac{\dfrac{\mathrm{Cov}(Y,Q)}{\mathrm{Var}(Y)}}{\dfrac{\mathrm{Cov}(Y,P)}{\mathrm{Var}(Y)}} = \frac{\mathrm{Cov}(Y,Q)}{\mathrm{Cov}(Y,P)} = \hat{\alpha}_2^{IV}$$

$$Q_{Dt} = \beta_1 + \beta_2 P_t + \beta_3 Y_t + u_{Dt} \qquad (1)$$
$$Q_{St} = \alpha_1 + \alpha_2 P_t + u_{St} \qquad (2)$$

At the first stage of TSLS we first run regression of P on available instrument Y:

$$\hat{P} = \hat{\gamma}_1 + \hat{\gamma}_2 Y$$

and then at the second stage use forecasted value of $\hat{P}$ as an explanatory variable in OLS instead of endogenous variable P: $\hat{\alpha}_2^{TSLS} = \dfrac{\mathrm{Cov}(\hat{P},Q)}{\mathrm{Var}(\hat{P})}$.

Let us show that it is possible to use at the second stage the IV estimator $\dfrac{\mathrm{Cov}(\hat{P},Q)}{\mathrm{Cov}(\hat{P},P)}$. It is sufficient to show that $\mathrm{Cov}(\hat{P},P) = \mathrm{Cov}(\hat{P},\hat{P}+e) = \mathrm{Cov}(\hat{P},\hat{P}) + \mathrm{Cov}(\hat{P},e) = \mathrm{Var}(\hat{P}) + 0 = \mathrm{Var}(\hat{P})$ according well known property of OLS $\mathrm{Cov}(\hat{P},e) = 0$.

Now

$$\hat{\alpha}_2^{TSLS} = \frac{\mathrm{Cov}(\hat{P},Q)}{\mathrm{Var}(\hat{P})} = \frac{\mathrm{Cov}(\hat{P},Q)}{\mathrm{Cov}(\hat{P},P)} = \frac{\mathrm{Cov}(P+e,Q)}{\mathrm{Cov}(P+e,P)} = \frac{\mathrm{Cov}(\hat{\gamma}_1+\hat{\gamma}_2 Y,Q)}{\mathrm{Cov}(\hat{\gamma}_1+\hat{\gamma}_2 Y,P)} = \frac{\mathrm{Cov}(\hat{\gamma}_1,Q)+\hat{\gamma}_2\,\mathrm{Cov}(Y,Q)}{\mathrm{Cov}(\hat{\gamma}_1,P)+\hat{\gamma}_2\,\mathrm{Cov}(Y,P)} =$$
$$\frac{0+\hat{\gamma}_2\,\mathrm{Cov}(Y,Q)}{0+\hat{\gamma}_2\,\mathrm{Cov}(Y,P)} = \frac{\hat{\gamma}_2\,\mathrm{Cov}(Y,Q)}{\hat{\gamma}_2\,\mathrm{Cov}(Y,P)} = \frac{\mathrm{Cov}(Y,Q)}{\mathrm{Cov}(Y,P)} = \hat{\alpha}_2^{IV}$$

So, $\hat{\alpha}_2^{ILS} = \hat{\alpha}_2^{IV} = \hat{\alpha}_2^{TSLS}$

5. [30 marks] To find factors influencing the employment of women, following estimates were calculated from a sample of 640 women respondents. The dependent variable EMP_i takes the value 1 if the woman was in paid employment and 0 otherwise.

(1) Method of estimation: OLS

$$P(EMP_i = 1) = -0.068 + 0.093UNIV_i - 0.210SCHOOL_i + 0.038AGE_i - 0.051AGE2_i + 0.024MARRIED_i$$

$$(0.049)\quad (0.015)\qquad (0.013)\qquad (0.003)\qquad (0.003)\qquad (0.009)\quad R^2 = 0.192$$

(2) Method of estimation: PROBIT)

$$P(EMP_i = 1) = -1.593 + 0.259UNIV_i - 0.554SCHOOL_i + 0.107AGE_i - 0.142AGE2_i + 0.063MARRIED_i$$

$$(0.137)\quad (0.043)\qquad (0.035)\qquad (0.008)\qquad (0.009)\qquad (0.035)$$

$$\ln L_U = -321.25 \quad \ln L_R = -416.01$$

Hint: in probit model the standard normal probability function is used $F(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z} e^{-\frac{1}{2}x^2} dx$.

where $UNIV_i$ is 1 if the respondent has a university qualification, 0 otherwise; $SCHOOL_i$ is 1 if the respondent graduated school but has no further professional or university qualification, 0 otherwise; AGE_i is age of a woman in years; $AGE2_i = AGE_i^2/100$; $MARRIED_i$ is 1 if married, 0 otherwise. Conventionally calculated standard errors are in brackets for the ordinary least squares (OLS) results, asymptotic standard errors are in brackets elsewhere. $\ln L_U$ is the log of the likelihood from the unrestricted probit model (as it presented in the equation above) and $\ln L_R$ is the log of the likelihood from the restricted probit model (the same model estimated with no explanatory variable included in the regression).

(a) [5 marks] Explain why in the models (1) and (2) evaluated by different methods, the coefficients turned out to be different in magnitude. Explain the logic of estimating of equation (2). Which of the two equations (1) or (2) can be trusted more and why?

Estimated values of dependent variable $\hat{EMP}_i$ can be interpreted as the probability for woman to be in paid emploiment. Equation (1) is an linear probability model (LPM) and is evaluated using OLS, while (2) is probit model, the logic of its estimation is more complicated. Values Z_i of the left side of equation (2) $Z = \beta_0 + \beta_1 UNIV + \beta_2 SCHOOL + \beta_3 AGE + \beta_4 AGE2 + \beta_5 MARRIED + u$ are used as the arguments for the transformation $F(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z} e^{-\frac{1}{2}x^2} dx$ that provides values of probability to be in paid employment $\hat{EMP}_i$.

For estimation of the parameters of this equation maximum likelihood method is used: likelihood function is $L = \frac{1}{\sqrt{2\pi}} \prod e^{-\frac{1}{2}x_i^2}$.

As the model (1) is linear in the variables the predicted values for it could be outside of the interval [0; 1]. In the probit model the curve $EMP = F(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z} e^{-\frac{1}{2}x^2} dx$ being the cumulative frequency of normal distribution lies in the interval [0; 1].

The second disadvantage of the model (1) is that marginal effects are constant. In reality for example the marginal effect of age on employment cannot be constant. In the probit model the curve $F(Z)$ has different slope in each point so the marginal effect changes.

The third disadvantage of LPM is heteroscedasticity that is inevitable in the case of using binary choice variable. For example most of observations corresponding to very young or very old women will have small residuals as actual state of unemployment will coincide with the prediction by the model, while for middle ages the situation with employment becomes more uncertain so many residuals could be relatively big. The heteroscedasticity violates GM conditions making OLS estimates inefficient. MLE is less sensitive to the violation of GMC.

The fourth disadvantage of LPM is connected with the distribution of disturbance term that is not normal when using binary dependent variable. All test in OLS become invalid as they are based on the assumption of normality of residuals. Under MLE we use different set of tests based on the likelihood function.

(b) [5 marks] How to interpret the coefficients of model (1)? What is the reference category for the set of dummy variables? Explain why the coefficients of $SCHOOL_i$ has negative sign. Test the significance of variables in equations (1) and (2)?

The coefficients of LPM (1) are marginal effects of different factors.
There are three categories for education: school as highest level of education, professional qualification and university level education. The reference categories are unmarried women with professional qualification, so the negative sign of the coefficient of $SCHOOL_i$ reflects the fact that women with school as highest level of education have less chance to be emploied than women with professional qualification.

In both models it is possible to test significance of all factors except age as it is represented by two variables AGE_i and $AGE2_i$. In LPM model one need to evaluate t-statistics $t = \frac{\hat{\beta}}{s.e.(\hat{\beta})}$ and then compare them with critical value from normal distribution (as the number of observation is large enough). The constant term is insignificant while $UNIV_i$, $SCHOOL_i$, and $MARRIED_i$ are significant at 1% level. Almost the same procedure is applicable to probit model, but we should have in mind that the equation uses standard errors based on the asymptotic z-distribution. The results are different: only $UNIV_i$ and $SCHOOL_i$ are significant at 1% level, while constant term and $MARRIED_i$ are insignificant (we cannot use one-sided test for $MARRIED_i$ because we are not sure in the direction of its influence).

(c) [6 marks] Why the value of R-squared is missing for equation (2)? What value can be used instead of R-squared? Are equations (1) and (2) as a whole significant? How to evaluate the significance of the variables AGE_i and $AGE2_i$, what should be done for that?

Determination coefficient for LPM is $R^2 = 0.192$, it is enough to make the equation (1) significant $F = \frac{R^2/5}{(1-R^2)/(640-6)} = \frac{0.192/5}{(1-0.192)/636} = 30.23$ what is much greater than $F_{crit}(5,\infty) = 3.02$.

For model (2) McFadden R^2 is $1 - \frac{\log L_U}{\log L_R} = 1 - \frac{321.25}{416.01} = 0.228$. It has nothing to do with the percentage of explained variance of dependent variable. The probit model is also significant as under H_0: All the slope coefficients are $=0$, $-2[\ln L_R - \ln L_U] \sim \chi^2_5$, $-2[-416.01-(-321.25)] = 189.92$. Critical value of χ^2_5 at 5% level of significance is 15.086, hence reject H_0.

To test the significance of age we should run additionally the restricted versions of equations (1) and (2) (without AGE_i and $AGE2_i$): $EMP_i = \beta_1 + \beta_2 UNIV_i + \beta_3 SCHOOL_i + \beta_4 MARRIED_i + w_i$ and memorize R^2_R for LPM model and $\log L_R$ for probit model, and then use F-test $\frac{(R^2_U - R^2_R)/2}{(1-R^2_U)/(640-6)}$ for LPM, and $\chi^2(2)$ test for probit model with LR statistic $LR = 2(\log L_R - \log L_U)$

(d) [7 marks] Calculate the estimated probabilities of being in employment for an unmarried woman with the university education at the age 40 and 45. Comment on your results. At what age the probability of being employed is at its maximum? It is natural to expect that for the women with university education this 'optimal' age is different. What variable(s) can be added in the model to take this into account?

LPM(40): =-0.068+0.093*1-0.210*0+0.038*40-0.051*40^2/100+0.024*0=0.729
LPM(45): =-0.068+0.093*1-0.210*0+0.038*45-0.051*45^2/100+0.024*0=0.702
Delta=0.702-0.729=-0.023 => decreases by 0.023 percentage points

PROBIT(40): Z=-1.593+0.259*1-0.554*0+0.107*40-0.142*40^2/100+0.062*0=0.674

Using Normal Tables find F(0.674)=0.750
PROBIT(45): Z=-1.593+0.259*1-0.554*0+0.107*45-0.142*45^2/100+0.062*0=0.6055
Using Normal Tables find F(0.6055)=0.727
Delta=0.727-0.75=-0.023, so the results are approximately similar.

The negative sign of $AGE2_i = AGE^2/100$ along with positive sign of AGE_i tells that that age increases the chance to be employed only for young women. When woman is getting older the negative effect of $AGE2_i = AGE^2/100$ begins to exceed positive effect of AGE_i and so the total effect becomes negative. There is an optimal age when the chances of woman to be emploied are the greatest, for LPM it is 37.7:

$$\frac{d(0.38AGE - 0.51AGE^2/100)}{dAGE} = 0.38 - 2\cdot 0.51AGE/100 = 0 \Rightarrow AGE = \frac{38}{1.02} = 37.25.$$

According probit model it is almost the same

$$\frac{d(0.107AGE - 0.142AGE^2/100)}{dAGE} = 0.107 - 2\cdot 0.142AGE/100 = 0 \Rightarrow AGE = \frac{10.7}{0.284} = 37.68$$

This result does not depend on the education level. The simplest way to take university education into account is to add intercaction dummy $UNIV_i * AGE_i$ to the model, run regression and then repeat the calculations of derivative.

(e) [7 marks] Give the interpretation to the coefficient of the variable MARRIED in model (2). Evaluate the marginal effect of marriage for an unmarried woman aged 20 having neither university nor professional qualification using OLS, and Probit estimates. Whether the estimated effect coincides with your expectations? Comment.

The coefficients of the model (2) do not show marginal effects of factors directly, as the structure of this model is more complicated. In probit model EMP is a function of Z which is in turn a function of some factors $X_1, X_2,...$: $EMP = F(Z(X_k))$, $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...$, so $\frac{dEMP}{dX_k} = \frac{dF}{dZ}\cdot\frac{\partial Z}{X_k} = \frac{dF}{dZ}\cdot\beta_k$.

Probit. To evaluate marginal effect of marriage one should multiply the coefficient of *MARRIED* by the deribative of $F(Z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{Z} e^{-\frac{1}{2}x^2}dx$:

$$\frac{\partial F(Z)}{\partial MARRIED} = \frac{dF(Z)}{dZ}\cdot\frac{\partial Z}{\partial MARRIED} = f(Z)\cdot\hat{\beta}_{MARRIED} = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}Z^2}\cdot 0.063$$

First let us evaluate *Z*=-1.593+0.259*0-0.554*1+0.107*20-0.142*20^2/100+0.063*0= - 0.575

So $f(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}\cdot 0.063 = 0.338\cdot 0.063 = 0.021$, this means that marriage for woman increases the probability of being employed by 2.1 percentage points.
As we have seen in (b) the marginal effect of marriage is insignificant.
Let us compare with LPM.
LPM. The marginal effect of marriage in LPM is constant and equal to 0.024, this means that marriage for woman increases the probability of being employed by 2.4 percentage points. The effect is significant.

Alternative way to do the same is to use direct calculation and comparison of probabilities of being employed for unmarried and married women.
Z(unmarried)=-1.593+0.259*0-0.554*1+0.107*20-0.142*20^2/100+0.063*0= - 0.575
Z(married)=-1.593+0.259*0-0.554*1+0.107*20-0.142*20^2/100+0.063*1= - 0.512
=-1.593+0.259*0-0.554*1+0.107*20-0.142*20^2/100+0.063*1=-0.512
For the Function F(Z): F(-Z)=1-F(Z), so
Delta= F(-0.512)- F(-0.575)=1-F(0.512)-1+ F(0.575)= F(0.575) -F(0.512)=0.717-0.696=0.021
The marriage rises the probability of being emploied. There could be many different possible explanation of this fact (really interesting because there are many obvious arguments in favour of the opposite effect). The most interesting fact is that probably men value in women the same qualities as the employers (this similarity of the marriage and emploiment was indicated in many econometrics papers).