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:

$$\underset{(s.e.)}{LOCT_t} = \underset{(1.02)}{5.31} - \underset{(0.15)}{0.28}\ln DPI_t$$

(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 $\frac{0.28}{100} = 0.0028$ billions of dollars $= 2{,}800$ thousands of dollars. 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 2 Discuss how dummy variables can be used to test

(a) change in intercept;

Unlike the quantitative variables the qualitative variables whose values are the categories cannot be used directly in regression models. To express the fact that observations refer to a certain category, it is necessary to use special auxiliary variables, called dummies. Suppose that there are only two categories, for example before and after certain event, e.g. crisis. Define dummy variable $D_t = 0$ for all observations before crisis ($t = 1, 2, ..., s$), and $D_t = 1$ after crisis ($t = s+1, s+2, ..., T$). Let equation $Y_t = \beta_1 + \beta_2 P_t + u_t$ be the dependence of the consumption of certain type of goods Y_t from the prices P_t of these goods. Inclusion of the dummy variable D_t into equation:

$$Y_t = \beta_1 + \beta_2 P_t + \beta_3 D_t + u_t$$

allows to measure the effect of choosing the period (before and after crisis) on the intercept, assuming common value of the slope:

$$Y_t = \beta_1 + \beta_2 P_t + u_t \text{ before crisis,}$$

$$Y_t = (\beta_1 + \beta_3) + \beta_2 P_t + u_t \text{ after crisis.}$$

To measure the significance of this effect it is sufficient to test the significance of β_3 using t-test (alternatively F-test for restriction $\beta_3 = 0$ in equation $Y_t = \beta_1 + \beta_2 P_t + \beta_3 D_t + u_t$ could be used).

(b) change in slope;

if the goal is to study the effect of a category change on the slope of the equation, then the variable introduced above should be included in the equation in another way. First, we determine the slope dummy by simply multiplying the intercept dummy by the quantitative explanatory variable, and then include it into equation as follows:

$$Y_t = \beta_1 + \beta_2 P_t + \beta_3 D_t P_t + u_t$$

So we have

$$Y_t = \beta_1 + \beta_2 P_t + u_t \text{ before crisis,}$$

$$Y_t = \beta_1 + (\beta_2 + \beta_3) P_t + u_t \text{ after crisis.}$$

(we assume here that intercept is common for both periods, alternatively one can consider the following equation $Y_t = \beta_1 + \beta_2 P_t + \beta_3 D_t P_t + \beta_4 D_t + u_t$). To measure the significance of this effect it is sufficient to test the significance of β_3 using t-test (alternatively F-test for restriction $\beta_3 = 0$ in equation $Y_t = \beta_1 + \beta_2 P_t + \beta_3 D_t P_t + u_t$ could be used).

(c) changes in both intercept and slope.

To take into account the influence of changing category (before crisis, after crisis) on both intercept and slope we should consider equation

$$Y_t = \beta_1 + \beta_2 D_t + \beta_3 P_t + \beta_4 D_t P_t + u_t$$

For the period before crisis it looks like

$$Y_t = \beta_1 + \beta_3 P_t + u_t$$

For the period after crisis it looks like

$$Y_t = (\beta_1 + \beta_2) + (\beta_3 + \beta_4) P_t + u_t$$

To measure this joint effect F-test for two restrictions $\beta_2 = 0,\ \beta_4 = 0$ can be used. To do this test practically first run unrestricted model $Y_t = \beta_1 + \beta_2 D_t + \beta_3 P_t + \beta_4 D_t P_t + u_t$ using all sample $t = 1, 2, ..., T$ and estimate residual sum of squares RSS_U. Then run resricted model $Y_t = \beta_1 + \beta_3 P_t + u_t$ using the same sample, and estimate residual sum of squares RSS_R. Now we can evaluate F-statistic $F=\frac{(RSS_R-RSS_U)/2}{RSS_U/(T-4)}$ having F-distribution with $(2,T-4)$ degrees of freedom.

Alternatively one can use Chow test: using original regression without dummies

$$Y_t=\beta_1+\beta_2P_t+u_t$$

first run it for the total sample $t=1,2,...,T$ and evaluate RSS_{Total}. Then run this regression in turn for the period before crisis $t=1,2,...,s$ evaluating RSS_{Before}, and the for the period after crisis $t=s+1,s+2,...,T$, evaluating RSS_{After}; now it is possible to evaluate F-statistic $F=\frac{(RSS_{Total}-(RSS_{Before}+RSS_{After})/2}{(RSS_{Before}+RSS_{After})/(T-2\cdot 2)}$. As $RSS_R=RSS_{Total}$ and $RSS_U=RSS_{Before}+RSS_{After}$ F-statistic for Chow test is the same as F-statistic for joint significance of two dummies, their critical values also coincide, so they are equivalent.

Question 3 The simple linear regression $Y_i=\beta_1+\beta_2X_i+u_i$, $i=1,2,...,n$ is considered. $\hat{Y}_i=b_1+b_2X_i$ are estimated values of dependent variable, the residuals are $\hat{u}_i=Y_i-b_1-b_2X_i$ where b_1 and b_2 are ordinary least squares estimates of β_1 and β_2.

(a) Show that $\text{Var}(Y_i)=\text{Var}(\hat{Y}_i)+\text{Var}(\hat{u}_i)$, explaining clearly how this equality follows from OLS principle.

$Y_i=\hat{Y}_i+\hat{u}_i$ hence $\text{Var}(Y_i)=\text{Var}(\hat{Y}_i)+\text{Var}(\hat{u}_i)+2\,\text{Cov}(\hat{Y}_i,\hat{u}_i)$. Now,

$\text{Cov}(\hat{Y}_i,\hat{u}_i)=\text{Cov}(b_1+b_2X_i,\hat{u}_i)=\text{Cov}(b_1,\hat{u}_i)+b_2\,\text{Cov}(X_i,\hat{u}_i)$, which is 0 since $\text{Cov}(b_1,\hat{u}_i)=0$ as b_1 is a constant and $\text{Cov}(X_i,\hat{u}_i)=0$ by the normal equations of OLS (see below *).

Hence $\text{Var}(Y_i)=\text{Var}(\hat{Y}_i)+\text{Var}(\hat{u}_i)$.

[Note: Variances and covariances are sample variances and sample covariances.]

*) By definition $\text{Cov}(X_i,\hat{u}_i)=\frac{1}{n}\left(\sum X_i\cdot\hat{u}_i\right)-\frac{1}{n}\left(\sum X_i\right)\cdot\frac{1}{n}\left(\sum\hat{u}_i\right)=\frac{1}{n}\left(\sum X_i\hat{u}_i\right)$ as $\sum\hat{u}_i=0$, so

$\text{Cov}(X_i,\hat{u}_i)=\frac{1}{n}(\sum X_i\hat{u}_i)=\frac{1}{n}(\sum X_i\cdot(Y_i-b_1-b_2X_i))=0$ as $\sum X_iY_i=b_1\sum X_i+b_2\sum X_iY_i$ 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 $\sum\hat{Y}_i\hat{u}_i=\sum(\hat{\beta}_1+\hat{\beta}_2X_i)\hat{u}_i=\hat{\beta}_1\sum\hat{u}_i+\hat{\beta}_2\sum X_i\hat{u}_i=0+0=0$

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

$\frac{1}{n}\sum(Y_i-\bar{Y})^2=\frac{1}{n}\sum(\hat{Y}_i-\bar{Y})^2+\frac{1}{n}\sum\hat{u}_i^2$ or $\text{Var}(Y_t)=\text{Var}(\hat{Y}_t)+\text{Var}(\hat{u}_t)$

(by definition $\text{Var}(X)=\frac{1}{n}\sum(X_i-\bar{X})^2$)

(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 $\mathrm{Cov}(Y_i, \hat{Y}_i) = \frac{1}{n}\sum(Y_i - \bar{Y})(\hat{Y}_i - \bar{Y})$ and $\mathrm{Var}(\hat{Y}_i) = \frac{1}{n}\sum(\hat{Y}_i - \bar{\hat{Y}})^2 = \frac{1}{n}\sum(\hat{Y}_i - \bar{Y})^2$ 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 $\sum\hat{Y}_i\hat{u}_i = \sum(\hat{\beta}_1 + \hat{\beta}_2 X_i)\hat{u}_i = \hat{\beta}_1\sum\hat{u}_i + \hat{\beta}_2\sum X_i\hat{u}_i = 0 + 0 = 0$

Alternative solution.

$$\mathrm{Cov}(Y_i, \hat{Y}_i) = \mathrm{Cov}((\hat{Y}_i + \hat{u}_i), \hat{Y}_i) = \mathrm{Cov}(\hat{Y}_i, \hat{Y}_i) + \mathrm{Cov}(\hat{u}_i, \hat{Y}_i) = \mathrm{Var}(\hat{Y}_i) + 0 = \mathrm{Var}(\hat{Y}_i)$$

as $\mathrm{Cov}(\hat{u}_i, \hat{Y}_i) = \frac{1}{n}\sum\hat{Y}_i\hat{u}_i = \frac{1}{n}\sum(b_1 + b_2X_i)\hat{u}_i = \frac{1}{n}(b_1\sum\hat{u}_i + b_2\sum X_i\hat{u}_i) = 0 + 0 = 0$

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.

$$r_{Y,\hat{Y}} = \frac{\mathrm{Cov}(Y_i, \hat{Y}_i)}{\sqrt{\mathrm{Var}(Y_i) \cdot \mathrm{Var}(\hat{Y}_i)}} = \frac{\mathrm{Var}(\hat{Y}_i)}{\sqrt{\mathrm{Var}(Y_i) \cdot \mathrm{Var}(\hat{Y}_i)}} = \frac{\sqrt{\mathrm{Var}(\hat{Y}_i)}}{\sqrt{\mathrm{Var}(Y_i)}} = \sqrt{\frac{\mathrm{Var}(\hat{Y}_i)}{\mathrm{Var}(Y_i)}} = \sqrt{R^2}$$

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

$= 0$ 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 one year of education S_i (schooling) 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:

$$y_t = ax_t + u_t,\ t = 1, ..., T$$

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.

$$a = a^*,\ t = 1, ..., s$$
$$a = a^{**},\ t = s+1, ..., T$$

(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 $t = 1, 2, ..., s$ and $t = 1, 2, ..., s$.

(ii) First introduce dummy variable $d_t = 0$ for $t = 1, 2, ..., s$ and $d_t = 1$ for $t = s+1, ..., n$.

Specify the model as $y_t = (a + \lambda d_t)x_t + u_t = ax_t + \lambda(d_t x_t) + u_t$ for $t = 1, 2, ..., T$ 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^*.

Solution and marking

b) $\hat{a} = \frac{\sum x_t y_t}{\sum x_t^2} = \frac{\sum x_t(ax_t + \lambda d_t x_t + u_t)}{\sum x_t^2} = a\frac{\sum x_t x_t}{\sum x_t^2} + \lambda\frac{\sum d_t x_t x_t}{\sum x_t^2} + \frac{\sum x_t u_t}{\sum x_t^2} = a + \lambda\frac{\sum_{t=s+1}^{T} x_t^2}{\sum_{t=1}^{T} x_t^2} + \frac{\sum x_t u_t}{\sum x_t^2}$ (here the sum $\sum$ means $\sum_{t=1}^{T}$, unless otherwise indicated).

so $E(\hat{a}) = a + \lambda \frac{\sum_{t=s+1}^{T} x_t^2}{\sum_{t=1}^{T} x_t^2} + \frac{\sum x_t E(u_t)}{\sum x_t^2} = a + \lambda \frac{\sum_{t=s+1}^{T} x_t^2}{\sum_{t=1}^{T} x_t^2}$ since $E(u_t) = 0$. The bias is therefore $\lambda \frac{\sum_{t=s+1}^{T} x_t^2}{\sum_{t=1}^{n} x_t^2} = \frac{\lambda \sum d_t x_t^2}{\sum x_t^2}$.

(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 $t = 1, 2, ..., T$ and use a t-test for the estimate of λ: $t = \frac{\hat{\lambda}}{s.e.(\hat{\lambda})}$ and compare it with $t_{crit}(\alpha, T-2)$.

(ii) Estimate the model as $y_t = ax_t + u_t$ for $t = 1, 2, ..., T$ 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 $t = 1, 2, ..., T$ and memorize unrestricted sum of squares RSS_U. Now evaluate F-test for restriction $\lambda = 0$: $F_{1,T-2} = \frac{(RSS_R - RSS_U)/1}{RSS_U/(T-2)}$. 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 $F_{1,T-2} = \frac{(RSS_R - RSS_U)/1}{RSS_U/(T-2)}$. 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:

$$E(Y_t) = Y_0 + \sum_{s=1}^{t} Eu_s = Y_0 \text{ (constant in } t)$$

but:

$$\text{Var}(Y_t) = \text{Var}\left(\sum_{s=1}^{t} u_s\right) = t\sigma^2.$$

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:

$$E(u_t) = E(e_t + \theta e_{t-1}) = Ee_t + \theta E e_{t-1} = 0.$$
$$E(u_t^2) = E(e_t^2) + \theta^2 E(e_{t-1}^2) + 2E(e_t e_{t-1}) = (1+\theta^2)\sigma^2,$$

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).

$$\ln \hat{TOB}_t = \underset{(0.09)}{1.56} + \underset{(0.02)}{0.21} \ln PTOB_t \qquad R^2 = 0.82,\ d = 0.76 \qquad (1)$$

Trying to remove autocorrelation the student applies autoregressive transformation AR(1)

$$\ln \hat{TOB}_t = \underset{(4.84)}{5.18} - \underset{(0.20)}{0.28} \ln PTOB_t \qquad R^2 = 0.92,\ d = 2.61 \qquad (2)$$

(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):

$$\underset{}{\ln \widehat{TOB}_t} = \underset{(1.74)}{7.23} - \underset{(0.11)}{0.37} \ln PTOB_t + \underset{(0.22)}{0.55} \ln DPI_t + \underset{(0.012)}{0.049} TIME_t \qquad R^2 = 0.93, \; d = 1.26 \qquad (3)$$

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

$$\ln T\hat{O}B_t = 0.25 + 0.90 \ln TOB_{t-1} \qquad R^2 = 0.90 \qquad (4)$$
$$(0.16) \quad (0.06) \qquad\qquad d = 2.17$$

Did a friend's advice help get rid of autocorrelation?

One should use h-statistics ($h = (1 - d/2)\sqrt{\frac{n}{1 - n \cdot \mathrm{Var}(b)}}$) which is equal here $h = (1 - 2.17)\sqrt{\frac{24}{1 - 24 \cdot (0.06)^2}} = -0.435$. 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.