ISEG – Lisbon School of Economics and Management
ECONOMETRICS
First Semester 2017/2018
Problem Set II

Question:	**1**	**2**	**3**	**4**	**5**	Total
Points:	4	4	4	4	34	50

Justify all your answers (except for multiple choice questions). You are required to show your work on each problem (except for multiple choice questions). **Organize your work**. Work scattered all over the page will receive very little credit. A correct answer in a multiple choice question worths 4 points; an incorrect one worths -1 point. **Delivery date: 24 of October**.

(4) **1**. Suppose that assumption MLR.5 is not verified in the model. Which of the following statements is **TRUE**?

- √ **The *t* and *F* statistics, obtained from the conventional OLS estimator, do not follow the usual distributions.**
- ◯ The OLS standard errors are unbiased, once assumption MLR.4 still holds.
- ◯ The conditional variance of the error term is a constant, σ^2.
- ◯ The OLS estimator is BLUE, once assumption MLR.4 still holds.

(4) **2**. Assume that you had estimated the following quadratic regression model:

$$\widehat{wage} = 6.1 + 0.35\, educ + 0.25\, exper - 0.01\, exper^2$$

If experience increases from 5 to 6, holding education fixed, then it is estimated that wage increases, in average, approximately:

- ◯ 0.25
- ◯ 0.20
- ◯ 0.90
- √ **0.15**

(4) **3**. Suppose you are performing a test $H_0 : \beta_j = 0$ against $H_1 : \beta_j < 0$ and you obtain the observed value for the test statistic equal to t_{obs}. Then the p-value is equal to:

- ◯ The probability of rejection of the null hypothesis.
- √ $P(T \leq t_{obs})$ **with T distributed according to a *t-student* with $n-k-1$ degrees of freedom.**
- ◯ $P(T \leq -t_{obs})$ with T distributed according to a *t-student* with $n-k-1$ degrees of freedom.
- ◯ $2P(T \geq |t_{obs}|)$ with T distributed according to a *t-student* with $n-k-1$ degrees of freedom.

(4) **4.** Consider the regression model:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

We intend to test the hypothesis that if x_1 and x_2 increase by the same amount then the estimated effect on y is in average equal to zero. The null hypothesis will be:

- √ $H_0 : \beta_1 + \beta_2 = 0$
- ○ $H_0 : \beta_1 - \beta_2 = 0$
- ○ $H_0 : E(\beta_1 - \beta_2) = 0$
- ○ none of the above.

5. Use the data set affairs.WF1 to explain the numbers of affairs a married person had last year. Estimate the following regression by OLS:

$$naffairs_i = \beta_0 + \beta_1\, educ_i + \beta_2\, age_i + \beta_3\, yrsmarr_i + \beta_4\, yrsmarr_i^2 + \beta_5\, relig_i + u_i$$

where:

- ***naffairs*** is the number of affairs within last year;
- ***educ*** is number of years in schooling;
- ***yrsmarr*** is the number of years married;
- ***relig*** gives the religious status of the individual where 5 = very religious, 4 = somewhat, 3 = slightly, 2 = not at all, 1 = anti religious.

(5) (a) Write the estimated equation with the corresponding standard errors.

Solution:

Dependent Variable: NAFFAIRS
Method: Least Squares

Sample: 1 601
Included observations: 601

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.856264	1.059552	2.695729	0.0072
EDUC	-0.015659	0.055216	-0.283598	0.7768
AGE	-0.036179	0.023026	-1.571268	0.1167
YRSMARR	0.276098	0.110749	2.492998	0.0129
YRSMARR^2	-0.005574	0.006560	-0.849665	0.3959
RELIG	-0.540427	0.114341	-4.726437	0.0000

R-squared	0.077350	Mean dependent var	1.455907
Adjusted R-squared	0.069597	S.D. dependent var	3.298758
S.E. of regression	3.181896	Akaike info criterion	5.162765
Sum squared resid	6024.054	Schwarz criterion	5.206677
Log likelihood	-1545.411	Hannan-Quinn criter.	5.179858
F-statistic	9.976374	Durbin-Watson stat	1.870104
Prob(F-statistic)	0.000000		

$$\widehat{naffairs} = 2.8563 - 0.0157\, educ - 0.0362\, age + 0.2761\, yrsmarr - 0.0056\, yrsmarr^2 - 0.5404\, relig$$

$se(\hat{\beta}_1) = 0.0552$; $se(\hat{\beta}_2) = 0.0230$; $se(\hat{\beta}_3) = 0.1107$; $se(\hat{\beta}_4) = 0.0066$; $se(\hat{\beta}_5) = 0.1143$

(7) (b) Interpret the estimated coefficients $\hat{\beta}_3$ and $\hat{\beta}_5$. Discuss the signs of these estimates.

Solution:
$\hat{\beta}_3 = 0.2761$: Since the variable *yrsmarr* appears also in $yrsmarr^2$ in the model, $\hat{\beta}_3$ has a slightly different interpretation: it represents, *ceteris paribus*, the estimated average increase in number of affairs that a newly-married person has in the first year of marriage. ($yrsmarr = 0$).

- This sign could perfectly be the opposite one - the fact that a person is newly-wed could affect his/her number of affairs negatively.

$\hat{\beta}_5 = -0.5404$: Holding all other factors fixed, an alteration in the religion status of a person (by 1 value, upwards) has an estimated average effect of -0.5404 in the number of affairs in an year. Since we are not using dummy variables, we assume that this effect is constant for all changes (by 1 value) in the categories.

- This sign may make sense - a more religious person may feel less interested in starting an affair, due to moral beliefs.

(5) (c) Write the estimated marginal effect of ***yrsmarr*** on ***naffairs*** in the specified model and comment.

Solution:

$$\frac{\partial \widehat{naffairs}}{\partial yrsmarr} = 0.2761 + 2 \times (-0.0056)\, yrsmarr = 0.2761 - 0.0112\, yrsmarr$$

The marginal effect of *yrsmarr* on *naffairs* is at first positive, but it decreases with the advance of the marriage - after some years, it becomes negative.
To get the turning point, we just need to find *naffairs* such as the marginal effect equals 0:

$$0.2761 - 0.0112\, yrsmarr = 0 \Leftrightarrow yrsmarr = 24.65$$

It is in the 25th year of marriage that the advance of the marriage has a negative effect on the number of affairs.

(6) (d) Test the individual statistical significance of all variables and comment.

Solution:
For each coefficient, we perform the test:

$$\text{H0: } \beta_j = 0 \quad \text{vs} \quad \text{H1: } \beta_j \neq 0,\ j \in \{1,2,3,4,5\}$$

The statistic we use for testing is:

$$t_{\hat{\beta}_j} = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \sim t(n-k-1),\ j \in \{1,2,3,4,5\}$$

Where $n - k - 1 = 601 - 5 - 1 = 595$.
We can get the p-values for all tests directly in the Eviews output.

- For β_1, we get a p-value of 0.7768 - failing to reject the null hypothesis with a significance level of 5%;
- For β_2, we get a p-value of 0.1167 - failing to reject the null hypothesis at the 5% level;
- For β_3, we get a p-value of 0.0129 - rejecting the null hypothesis at the 5% level;
- For β_4, we get a p-value of 0.3959 - failing to reject the null hypothesis at the 5% level;
- For β_5, we get a p-value of 0.0000 - rejecting the null hypothesis at the 5% level.

The conclusion is that the variables *educ*, *age* and $yrsmarried^2$ are not statistically significant: this is specially relevant for this last variable, since it indicates that there is no significant quadratic effect for the variable *yrsmarr*. On the other hand, *yrsmarr* and *relig* are statistically significant to explain *naffairs*.

(5) (e) Test whether the coefficients on ***educ*** and ***age*** are jointly significantly. Show your calculations.

Solution:

$$\text{H0: } \beta_1 = \beta_2 = 0 \quad \text{vs} \quad \text{H1: } \beta_1 \neq 0 \vee \beta_2 \neq 0$$

We have 2 linear restrictions, meaning that we have to perform an F-test. The F-statistic is:

$$F = \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n-k-1)} \sim F(q, n-k-1)$$

OR

$$F = \frac{(R^2_{ur} - R^2_r)/q}{(1 - R^2_{ur})/(n-k-1)} \sim F(q, n-k-1)$$

Where $q = 2$ (number of linear restrictions) and $n - k - 1 = 595$.

We need to estimate the restricted model, which is:

$$naffairs_i = \alpha_0 + \alpha_1\, yrsmarr_i + \alpha_2\, yrsmarr_i^2 + \alpha_3\, relig_i + v_i$$

Dependent Variable: NAFFAIRS
Method: Least Squares

Sample: 1 601
Included observations: 601

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.754490	0.470570	3.728437	0.0002
YRSMARR	0.259308	0.110265	2.351687	0.0190
YRSMARR^2	-0.007371	0.006423	-1.147676	0.2516
RELIG	-0.544763	0.114169	-4.771545	0.0000

R-squared	0.073001	Mean dependent var	1.455907
Adjusted R-squared	0.068343	S.D. dependent var	3.298758
S.E. of regression	3.184039	Akaike info criterion	5.160812
Sum squared resid	6052.449	Schwarz criterion	5.190087
Log likelihood	-1546.824	Hannan-Quinn criter.	5.172207
F-statistic	15.67132	Durbin-Watson stat	1.873291
Prob(F-statistic)	0.000000		

The sum of square residuals of the restricted model is 6052.449. The SSR of the model without restrictions (the original model) is 6024.054. We get an F-ratio of:

$$F_{obs} = \frac{(6052.449 - 6024.054)/2}{6024.054/595} = 1.4$$

With the R-squared Form:

$$F_{obs} = \frac{(0.07735 - 0.073)/2}{(1 - 0.07735)/595} = 1.4$$

We need to get the critical value of an F(2,595). For a 5% significance level, the 95th percentile is 3.01: $F_{obs} < 3.01$, meaning that we fail to reject the null hypothesis: the variables *educ* and *age* are not jointly statistically significant at the 5% level.

Alternatively, we can perform the test directly with Eviews (Wald Test-Coefficient Restrictions).

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
F-statistic	1.402273	(2, 595)	0.2468
Chi-square	2.804546	2	0.2460

Null Hypothesis: C(2)=0, C(3)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(2)	-0.015659	0.055216
C(3)	-0.036179	0.023026

Restrictions are linear in coefficients.

Here, we get automatically that $F_{obs} = 1.4$, and with a p-value of 0.2468 we reach the same conclusion for a 5% significance level.

(6) (f) Estimate now the following regression by OLS:

$$naffairs_i = \beta_0 + \beta_1\, age_i + \beta_2\, yrsmarr_i + \beta_3\, relig_i + u_i$$

Estimate the effect on the number of affairs of an individual that is very religious and is 15 years married relatively to another that is 2 years married and anti-religious, having both the same age. Write the hypothesis that this effect is equal to zero and test it. Show how to calculate the test statistic.

Solution:

Dependent Variable: NAFFAIRS
Method: Least Squares

Sample: 1 601
Included observations: 601

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.929169	0.612907	4.779144	0.0000
AGE	-0.040800	0.022233	-1.835055	0.0670
YRSMARR	0.188286	0.037264	5.052705	0.0000
RELIG	-0.541489	0.113981	-4.750692	0.0000

R-squared	0.076167	Mean dependent var	1.455907
Adjusted R-squared	0.071525	S.D. dependent var	3.298758
S.E. of regression	3.178598	Akaike info criterion	5.157391
Sum squared resid	6031.780	Schwarz criterion	5.186666
Log likelihood	-1545.796	Hannan-Quinn criter.	5.168786
F-statistic	16.40694	Durbin-Watson stat	1.868978
Prob(F-statistic)	0.000000		

For the individual 1, $yrsmarr = 15$ and $relig = 5$; for the individual number 2, $yrsmarr = 2$ and $relig = 1$.
If the effect is the same, then:

$$\begin{aligned} & naffairs_2 = naffairs_1 \\ & \Leftrightarrow \beta_0 + \beta_1\, age_2 + \beta_2 \times 15 + \beta_3 \times 5 + u_2 = \beta_0 + \beta_1\, age_1 + \beta_2 \times 2 + \beta_3 \times 1 + u_1 \\ & \Leftrightarrow 15\,\beta_2 + 5\,\beta_3 = 2\,\beta_2 + \beta_3 \\ & \Leftrightarrow 13\,\beta_2 + 4\,\beta_3 = 0 \end{aligned}$$

This means that the etimated effect is:

$$13\,\hat{\beta}_2 + 4\,\hat{\beta}_3 = 13 \times 0.1883 + 4 \times -0.5415 = 0.2819$$

We need to test H0: $13\,\beta_2 + 4\,\beta_3 = 0$ vs H1: $13\,\beta_2 + 4\,\beta_3 \neq 0$. Since this is one linear restriction, we can use the t statistic:

$$t = \frac{13\,\hat{\beta}_2 + 4\,\hat{\beta}_3}{se(13\,\hat{\beta}_2 + 4\,\hat{\beta}_3)} \sim t(597)$$

$$\begin{aligned} se(13\,\hat{\beta}_2 + 4\,\hat{\beta}_3) &= \sqrt{13^2\,Var(\hat{\beta}_2) + 4^2\,Var(\hat{\beta}_3) + 2 \times 13 \times 4\,Cov(\hat{\beta}_2, \hat{\beta}_3)} \\ &= \sqrt{169 \times 0.0373^2 + 16 \times 0.1140^2 + 104 \times (-0.000465)} = 0.6283 \end{aligned}$$

Which means that:

$$t = \frac{0.2819}{0.6283} = 0.4487$$

For a significance level of 5% (in a bilateral test), the critical value for a t distribution with 594 degrees of freedom is 1.96. Since $|t| < 1.96$, we fail to reject the null hypothesis: there is statistical evidence that this effect is equal to zero.

Another way to test this restriction is to reparametrize the model. Considering:

$$13\,\beta_2 + 4\,\beta_3 = \theta \Leftrightarrow \beta_2 = \frac{1}{13}\theta - \frac{4}{13}\beta_3$$

Plugging this into the original model, we get:

$$\begin{aligned} naffairs &= \beta_0 + \beta_1\,age + \left(\frac{1}{13}\theta - \frac{4}{13}\beta_3\right)\,yrsmarr + \beta_3\,relig + u \\ &= \beta_0 + \beta_1\,age + \theta \times \frac{1}{13}\,yrsmarr + \beta_3\left(relig - \frac{4}{13}\,yrsmarr\right) + u \end{aligned}$$

Dependent Variable: NAFFAIRS
Method: Least Squares
Date: 10/19/17 Time: 23:59
Sample: 1 601
Included observations: 601

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.929169	0.612907	4.779144	0.0000
AGE	-0.040800	0.022233	-1.835055	0.0670
(1/13)*YRSMARR	0.281767	0.627813	0.448806	0.6537
RELIG-(4/13)*YRSMAR...	-0.541489	0.113981	-4.750692	0.0000

R-squared	0.076167	Mean dependent var	1.455907
Adjusted R-squared	0.071525	S.D. dependent var	3.298758
S.E. of regression	3.178598	Akaike info criterion	5.157391
Sum squared resid	6031.780	Schwarz criterion	5.186666
Log likelihood	-1545.796	Hannan-Quinn criter.	5.168786
F-statistic	16.40694	Durbin-Watson stat	1.868978
Prob(F-statistic)	0.000000		

Now we just have to test the hypothesis:

$$\text{H0: } \theta = 0 \quad \text{vs} \quad \text{H1: } \theta \neq 0$$

$$t_{\hat{\theta}} = \frac{\hat{\theta}}{se(\hat{\theta})} \sim t(597)$$

Using the Eviews output, we can easily see that the p-value for this test is 0.6537, which leads us to the conclusions taken before.

Again, we can use Eviews to test the restriction directly:

Wald Test:
Equation: Untitled

Test Statistic	Value	df	Probability
t-statistic	0.448806	597	0.6537
F-statistic	0.201427	(1, 597)	0.6537
Chi-square	0.201427	1	0.6536

Null Hypothesis: 13*C(3)+4*C(4)=0
Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
13*C(3) + 4*C(4)	0.281767	0.627813

Restrictions are linear in coefficients.

The p-value is (forcefully) the same as seen before - again, we should not reject the null hypothesis.