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

Question:	**1**	**2**	**3**	**4**	Total
Points:	4	4	4	38	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) and to include the output of EVIEWS used to solve the empirical 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: 30th of November**.

(4) **1**. Suppose the model $y = \beta_0 + \beta_1 x_1 + u$ where $Var(u|x_1 = \sigma^2 x_1^2)$ and $E(u|x_1) = 0$. Suppose also the models:

$$y/x_1 = \alpha_0 \times 1/x_1 + \alpha_1 + u/x_1$$

$$y/x_1^2 = \gamma_0 \times 1/x_1^2 + \gamma_1 + u/x_1^2$$

Which of the following statements is **TRUE**?

- ◯ The OLS estimator of β_0 and β_1 is BLUE.
- √ **The OLS estimator of α_0 and α_1 is BLUE.**
- ◯ The OLS estimator of γ_0 and γ_1 is BLUE.
- ◯ None of the above.

(4) **2**. Suppose the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$. Consider that $\hat{u}$ and $\hat{y}$ are the residuals and the fitted values for y obtained from estimating that model by OLS, respectively. Then the equation,

- ◯ $u^2 = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + v$ is used to perform the test of Breusch - Pagan.
- ◯ $\hat{u} = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + v$ is used to test for heteroscedasticity.
- ◯ $u^2 = \gamma_0 + \gamma_1 \hat{y} + \gamma_2 \hat{y}^2 + v$ is used to perform the RESET test.
- √ $\hat{u}^2 = \gamma_0 + \gamma_1 x_1 + +\gamma_2 x_1^2 + \gamma_3 x_1 x_2 + \gamma_4 x_2 + +\gamma_5 x_2^2 + v$ **is used to perform the test of White.**

(4) **3**. Choose the option that is FALSE. Suppose Assumptions MLR.1 to MLR.4 are valid. Then, the estimator of White for the standard errors in a multiple linear regression model,

- ◯ gives valid estimates with homoscedasticity and heteroscedasticity.
- ◯ should be used when there is evidence of heteroscedasticity.
- √ **gives valid estimates only for heteroscedasticity of the White type.**
- ◯ used in the t-statistic gives a statistic that is approximately normally distributed.

4. Use the data set mroz.WF1 to explain the numbers of hours a woman has worked in a given year.

Estimate the following regression by OLS:

$$hours_i = \beta_0 + \beta_1\, educ_i + \beta_2\, age_i + \beta_3\, log(faminc_i) + \beta_4\, kidslt6 + u_i$$

where:

- ***hours*** is the number of hours worked;
- ***educ*** is number of years in schooling;
- ***age*** is the woman's age in years;
- ***faminc*** is the family income;
- ***kidslt6*** is the number of kids with age less than 6 in the woman's household.

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

Solution:

Dependent Variable: HOURS
Method: Least Squares

Sample: 1 753
Included observations: 753

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1511.456	604.5454	-2.500153	0.0126
EDUC	20.62875	14.44602	1.427989	0.1537
AGE	-17.40978	4.159897	-4.185147	0.0000
LOG(FAMINC)	287.5780	63.32095	4.541593	0.0000
KIDSLT6	-478.7833	63.98363	-7.482903	0.0000

R-squared	0.108832	Mean dependent var	740.5764
Adjusted R-squared	0.104066	S.D. dependent var	871.3142
S.E. of regression	824.7320	Akaike info criterion	16.27461
Sum squared resid	5.09E+08	Schwarz criterion	16.30532
Log likelihood	-6122.391	Hannan-Quinn criter.	16.28644
F-statistic	22.83686	Durbin-Watson stat	1.131273
Prob(F-statistic)	0.000000		

$$\widehat{hours} = -1511.46 + 20.63\, educ - 17.41\, age + 287.58\, log(faminc) - 478.78\, kidslt6$$

$$se(\hat\beta_1) = 14.45;\ se(\hat\beta_2) = 4.16;\ se(\hat\beta_3) = 63.32;\ se(\hat\beta_4) = 63.98$$

(5) (b) Interpret the estimated coefficient $\hat\beta_3$ and discuss the signs of all the coefficient estimates.

Solution: $\hat{\beta}_3$: a raise of 1% in the family income will increase the estimated number of hours a woman works in an year by $\frac{287.58}{100} = 2.8758$, ceteris paribus.
All the signs of the estimates seem to make sense:

- A woman with more education probably has a more fulfilling job and doesn't mind working more hours;
- An older woman may feel less disposition to work several hours;
- A woman that receives more money may feel more motivated to work more (we are keeping all other factors constant);
- A woman with small kids may work less to spend more time with her children, who demand more attention.

(5) (c) Test for heteroscedasticity using the Breusch-Pagan test and conclude.

Solution: In all heteroscedasticity tests, the null hypothesis is the ausence of heteroscedastic errors.
In the Breusch-Pagan test, we have to perform the auxiliar regression:

$$\hat{u}^2 = \beta_0 + \beta_1\, educ + \beta_2\, age + \beta_3\, log(faminc) + \beta_4\, kidslt6 + v$$

Where $\hat{u}^2$ are the residuals of our original model.

$$\text{H0: } Var(u_i \mid X_i) = \sigma^2 \text{ vs}$$
$$\text{H1:} Var(u \mid X_i) = \gamma_0 + \gamma_1\, educ + \gamma_2\, age + \gamma_3\, log(faminc) + \gamma_4\, kidslt6$$

The null hypothesis can then be stated as:

$$\text{H0': } \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = 0$$

The test-statistic is

$$LM = nR^2_{\hat{u}^2} \xrightarrow{d} \chi^2(k)$$

Alternatively, we can perform the F-test for overall significance of the regression on squared residuals:

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

Dependent Variable: RESID^2
Method: Least Squares

Sample: 1 753
Included observations: 753

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1502351.	813826.7	-1.846033	0.0653
EDUC	-6172.435	19446.94	-0.317399	0.7510
AGE	-1309.733	5599.969	-0.233882	0.8151
LOG(FAMINC)	236964.8	85241.37	2.779927	0.0056
KIDSLT6	-173621.8	86133.46	-2.015729	0.0442

R-squared	0.018295	Mean dependent var	675666.3
Adjusted R-squared	0.013046	S.D. dependent var	1117551.
S.E. of regression	1110237.	Akaike info criterion	30.68466
Sum squared resid	9.22E+14	Schwarz criterion	30.71537
Log likelihood	-11547.78	Hannan-Quinn criter.	30.69649
F-statistic	3.484982	Durbin-Watson stat	2.053571
Prob(F-statistic)	0.007858		

For this situation, $nR^2_{\hat{u}^2} = 753 \times 0.018295 = 13.776135$

Considering $\alpha = 5\%$, the critical value for a chi-squared distribution with 4 degrees of freedom is 9.49: thus, we reject the null hypothesis, finding evidence that the errors are heteroscedastic.

Alternatively, we could use the F-statistic in the output of the regression which is equal to 3.48 with p-value 0.008. The same conclusion applies.

We may also make use of Eviews (View/Residual Diagnostics/Heteroskedasticity Tests) to get the result directly.

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	3.484982	Prob. F(4,748)	0.0079
Obs*R-squared	13.77637	Prob. Chi-Square(4)	0.0080
Scaled explained SS	18.56996	Prob. Chi-Square(4)	0.0010

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares

Sample: 1 753
Included observations: 753

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1502351.	813826.7	-1.846033	0.0653
EDUC	-6172.435	19446.94	-0.317399	0.7510
AGE	-1309.733	5599.969	-0.233882	0.8151
LOG(FAMINC)	236964.8	85241.37	2.779927	0.0056
KIDSLT6	-173621.8	86133.46	-2.015729	0.0442

R-squared	0.018295	Mean dependent var	675666.3
Adjusted R-squared	0.013046	S.D. dependent var	1117551.
S.E. of regression	1110237.	Akaike info criterion	30.68466
Sum squared resid	9.22E+14	Schwarz criterion	30.71537
Log likelihood	-11547.78	Hannan-Quinn criter.	30.69649
F-statistic	3.484982	Durbin-Watson stat	2.053571
Prob(F-statistic)	0.007858		

Using the first or the second line in the output, we get immediately that the p-value is smaller than 5%, reaching the same conclusion.

(5) (d) Test for heteroscedasticity using the White test and conclude.

Solution:

The White test, like the Breusch-Pagan, uses an auxiliar regression with the squared residuals, adding also the squares of the variables and their interactions:

$$\hat{u}^2 = \gamma_0 + \gamma_1\, educ + \gamma_2\, age + \gamma_3\, log(faminc) + \gamma_4\, kidslt6 + \gamma_5\, educ^2 + \gamma_6\, age^2 + \gamma_7\, log(faminc)^2 + \gamma_8\, kidslt6^2 + \gamma_9\, educ \times age + \gamma_{10}\, educ \times log(faminc) + \gamma_{11}\, educ \times kidslt6 + \gamma_{12}\, age \times log(faminc) + \gamma_{13}\, age \times kidslt6 + \gamma_{14}\, log(faminc) \times kidslt6 + v$$

Dependent Variable: RESID^2
Method: Least Squares

Sample: 1 753
Included observations: 753

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6262557.	9443544.	-0.663158	0.5074
EDUC	-55472.80	374702.1	-0.148045	0.8823
AGE	-106822.3	129133.5	-0.827223	0.4084
LOG(FAMINC)	1802846.	1667635.	1.081079	0.2800
KIDSLT6	-3114703.	1699909.	-1.832277	0.0673
EDUC^2	-4971.026	5891.239	-0.843800	0.3991
AGE^2	510.3236	724.0256	0.704842	0.4811
LOG(FAMINC)^2	-86372.66	89357.36	-0.966598	0.3341
KIDSLT6^2	28169.13	117404.4	0.239932	0.8104
EDUC*AGE	3619.592	2797.332	1.293944	0.1961
EDUC*LOG(FAMINC)	817.9793	40444.41	0.020225	0.9839
EDUC*KIDSLT6	37855.17	41068.38	0.921760	0.3570
AGE*LOG(FAMINC)	1615.044	12749.78	0.126672	0.8992
AGE*KIDSLT6	17406.58	18386.86	0.946686	0.3441
LOG(FAMINC)*KIDSLT6	182509.1	164430.5	1.109947	0.2674

R-squared	0.029328	Mean dependent var	675666.3
Adjusted R-squared	0.010915	S.D. dependent var	1117551.
S.E. of regression	1111435.	Akaike info criterion	30.69992
Sum squared resid	9.12E+14	Schwarz criterion	30.79203
Log likelihood	-11543.52	Hannan-Quinn criter.	30.73541
F-statistic	1.592737	Durbin-Watson stat	2.032834
Prob(F-statistic)	0.075665		

$$\text{H0: } Var(u_i \mid X_i) = \sigma^2 \text{ vs } \quad \text{H1: not H0}$$

The test-statistic is

$$LM = nR^2_{\hat{u}^2} \xrightarrow{d} \chi^2(q)$$

Or:

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

In this case, $nR^2_{\hat{u}^2} = 753 \times 0.029328 = 22.0842$

The critical value for a chi-squared distribution with 14 degrees of freedom is 23.7 ($\alpha = 5\%$) - we fail to reject the null hypothesis, concluding that there is evidence in favour of heteroscedastic errors when using the White Test.
Alternatively, more simple, the F-statistic is 1.59 with p-value 0.076 therefore we fail to reject H_0 at 5% leading to the same conclusion.

Using Eviews (View/Residual Diagnostics/Heteroskedasticity Tests):

Heteroskedasticity Test: White

F-statistic	1.592737	Prob. F(14,738)	0.0757
Obs*R-squared	22.08426	Prob. Chi-Square(14)	0.0769
Scaled explained SS	29.76864	Prob. Chi-Square(14)	0.0082

The conclusion is exactly the same.

(6) (e) Test for heteroscedasticity using the Simplified White test and conclude.

Solution: The Simplified White Test is used to conserve a small number of degrees of freedom. It is based on a regression that uses the fitted values of our dependent variable:

$$\hat{u}^2 = \alpha_0 + \alpha_1 \widehat{hours} + \alpha_2 \widehat{hours}^2 + v$$

$$\text{H0: } Var(u_i \mid X_i) = \sigma^2 \text{ vs } \quad \text{H1: not H0}$$

Dependent Variable: RESID^2
Method: Least Squares

Sample: 1 753
Included observations: 753

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	204758.8	144645.8	1.415587	0.1573
HOURSFIT	1024.879	417.9801	2.451980	0.0144
HOURSFIT^2	-0.456590	0.316528	-1.442496	0.1496

R-squared	0.016556	Mean dependent var	675666.3
Adjusted R-squared	0.013933	S.D. dependent var	1117551.
S.E. of regression	1109738.	Akaike info criterion	30.68112
Sum squared resid	9.24E+14	Schwarz criterion	30.69954
Log likelihood	-11548.44	Hannan-Quinn criter.	30.68822
F-statistic	6.312894	Durbin-Watson stat	2.050011
Prob(F-statistic)	0.001911		

The test-statistic is, once again,

$$LM = nR^2_{\hat{u}^2} \xrightarrow{d} \chi^2(2)$$

Or:

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

For this test, $nR^2_{\hat{u}^2} = 753 \times 0.016556 = 12.4667$

The critical value for a chi-squared distribution with 2 degrees of freedom is 5.99 ($\alpha = 5\%$) - we reject the null hypothesis, concluding that there is statistical evidence in favour of heteroscedastic errors.

Alternatively, more simple, the F-statistic is 6.31 with p-value 0.002 therefore we reject H_0 leading to the same conclusion.

(5) (f) Estimate the model using the White estimator for the standard errors.

Solution:

Dependent Variable: HOURS
Method: Least Squares

Sample: 1 753
Included observations: 753
White-Hinkley (HC1) heteroskedasticity consistent standard errors and covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1511.456	541.6069	-2.790688	0.0054
EDUC	20.62875	13.59506	1.517372	0.1296
AGE	-17.40978	4.313860	-4.035778	0.0001
LOG(FAMINC)	287.5780	55.81193	5.152626	0.0000
KIDSLT6	-478.7833	57.12056	-8.381978	0.0000

R-squared	0.108832	Mean dependent var	740.5764
Adjusted R-squared	0.104066	S.D. dependent var	871.3142
S.E. of regression	824.7320	Akaike info criterion	16.27461
Sum squared resid	5.09E+08	Schwarz criterion	16.30532
Log likelihood	-6122.391	Hannan-Quinn criter.	16.28644
F-statistic	22.83686	Durbin-Watson stat	1.131273
Prob(F-statistic)	0.000000	Wald F-statistic	33.30191
Prob(Wald F-statistic)	0.000000		

$$\widehat{hours} = -1511.46 + 20.63\, educ - 17.41\, age + 287.58\, log(faminc) - 478.78\, kidslt6$$

$$se(\hat{\beta}_1) = 13.60;\ se(\hat{\beta}_2) = 4.31;\ se(\hat{\beta}_3) = 55.81;\ se(\hat{\beta}_4) = 57.12$$

(7) (g) Given the results you obtained discuss the properties of the estimations in (a) and in (f).

Solution: Since there is evidence of heteroscedasticity with the first and third tests (and H0 in the White Test is not very far from being rejected), and supposing MRL.1 to MRL.4 apply, we should not use the usual standard errors computed in (a): all our inference (t-tests, F-tests, etc) will be invalid. Still, the OLS estimator for β is unbiased and consistent, even if it is not BLUE anymore.

Using OLS with White standard errors, as done in (f), allows us to perform correct inference, even if the distributions are only valid asymptotically.