Econometrics exam 07.02.2013

1. Exam takes 90 min.
2. This exam is a closed book exam.
3. Everybody is required to sign on the list.
4. The solution of exercise should be written on the sheet on which the exercise was printed or on the additional sheets on the back of the exam.
5. All the pages with solutions should be signed. If additional sheet is used it is very important to put the number of the exercise on the top of it.
6. Only one exercise should be solved on one sheet.
7. The minimum to obtain the pass grade is to answer two theoretical questions and to solve one exercise.

Theoretical questions

1	2	$\sum$

1. Why dicrete explanatory variable should be recoded into appropriate number of dummy variables before being included in the regression equation?
2. Derive OLS estimator for model with multiple explanatory variables. Show that the necessary and sufficient conditions for existence of the minimum of the objective function are in this case satisfied

Theoretical questions cont.

3	4	$\sum$

3. Explain the difference between parameters and estimates of the parameters and between error terms and residuals.

4. Why R^2 should not be used to compare models.

1	2	3	$\sum$

EXERCISE 1 The following model with one explanatory variable and no constant was analyzed:

$$y_t = \beta_i x_i + \varepsilon_i, \qquad \varepsilon \sim N\left(0, \sigma^2 \boldsymbol{I}\right)$$

where x_i is nonrandom. Number of observations $N = 11$, $\sum_{i=1}^{n} x_i^2 = 2$, $\sum_{i=1}^{n} x_i y_i = 4$ and the sum of residuals $\sum_{i=1}^{n} e_i^2 = 5$.

1. Calculate the value of the OLS estimator b of the parameter β.
2. Calculate the value of OLS estimator of variance of b.
3. It was suggested that the estimator $\widehat{\beta} = \frac{\overline{y}}{\overline{x}}$ is a good estimator of parameter β. Show that indeed this estimator is unbiased estimator of parameter β.

1	2	3	4	5	6	7	8a	8b	8c	$\sum$

EXERCISE 2 Wage equation for small enterprises was estimated on the data from Polish Labor Force Survey (LFS) from year 2010. Dependent variable (ln_wage) is the logarithm of wage. Explanatory variables are as follows: age (age), age squared (age_2), type of neighborhood in which employed is living ($rural$: 0 urban area, 1 rural area), sex (sex: 0 male, 1 female), education ($educ$: 0 primary, 1 vocational education, 2 secondary, 3 higher), working time (w_time: 0 part-time, 1 full-time), marital status ($status$: 0 single, 1 married, 2 widow/widower, 3 divorced) and interaction between gender and working time. Estimation results are reported below. Significance level to be used in testing $\alpha = 0.05$. Results of the tests should be justified with respective p-values.

```
      Source |       SS       df       MS              Number of obs =    1967
-------------+------------------------------           F( 12,  1954) =  118.05
       Model |  211.507908    12   17.625659           Prob > F      =  0.0000
    Residual |  291.740412  1954  .149304203           R-squared     =  0.4203
-------------+------------------------------           Adj R-squared =  0.4167
       Total |   503.24832  1966  .255975748           Root MSE      =   .3864
```

ln_wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ_1	.2021485	.0303484	6.66	0.000	.1426299	.2616671
educ_2	.2945919	.0362894	8.12	0.000	.2234219	.3657619
educ_3	.4560558	.0378275	12.06	0.000	.3818694	.5302422
rural	-.0433189	.0180795	-2.40	0.017	-.078776	-.0078618
age	.0388162	.0056042	6.93	0.000	.0278253	.0498072
age_2	-.0004376	.0000676	-6.47	0.000	-.0005702	-.0003049
sex	-.1072408	.0505662	-2.12	0.034	-.2064101	-.0080716
w_time	.780341	.0444233	17.57	0.000	.693219	.8674629
sexXw_time	-.1257776	.0534618	-2.35	0.019	-.2306257	-.0209296
status_1	.0951863	.0240566	3.96	0.000	.0480071	.1423656
status_2	-.1416434	.0605747	-2.34	0.019	-.2604411	-.0228456
status_3	-.0243112	.0445971	-0.55	0.586	-.111774	.0631517
_cons	5.470661	.106983	51.14	0.000	5.260849	5.680474

```
Breusch-Pagan LM statistic:       Chi2(1)    = 284.26  p-value = 0.000
White's general test statistic:   Chi2(70)   = 508.28  p-value = 0.000
Jarque-Berra test statistic:      Chi2(2)    = 3185.88 p-value = 0.000
Ramsey RESET test statistic:      F(3,1951)  = 44.72   p-value = 0.000
```

1. Are the explanatory variables jointly significant?
2. Give an interpretation to coefficient of determination R^2.
3. Verify which explanatory variables are statistically significant.
4. Interpret the values of the estimates for variables w_time and $sexXw\_time$. Use both the approximate and exact measure of the impact of the dummy explanatory variable. Calculate the partial semielasticity $wage$ w.r.t. age for a employed who are 40 years old.
5. Is the heteroscedasticity problem present in the estimated model?
6. Is the error term normal in the estimated model?
7. Check whether the linear functional form of the model is valid.
8. If estimated model does not satisfy the assumptions of CLRM explain:
   (a) which assumptions of CLRM are invalid,
   (b) what consequences does it have on statistical inference for this model,
   (c) how the problems indicated by diagnostic tests could be solved.

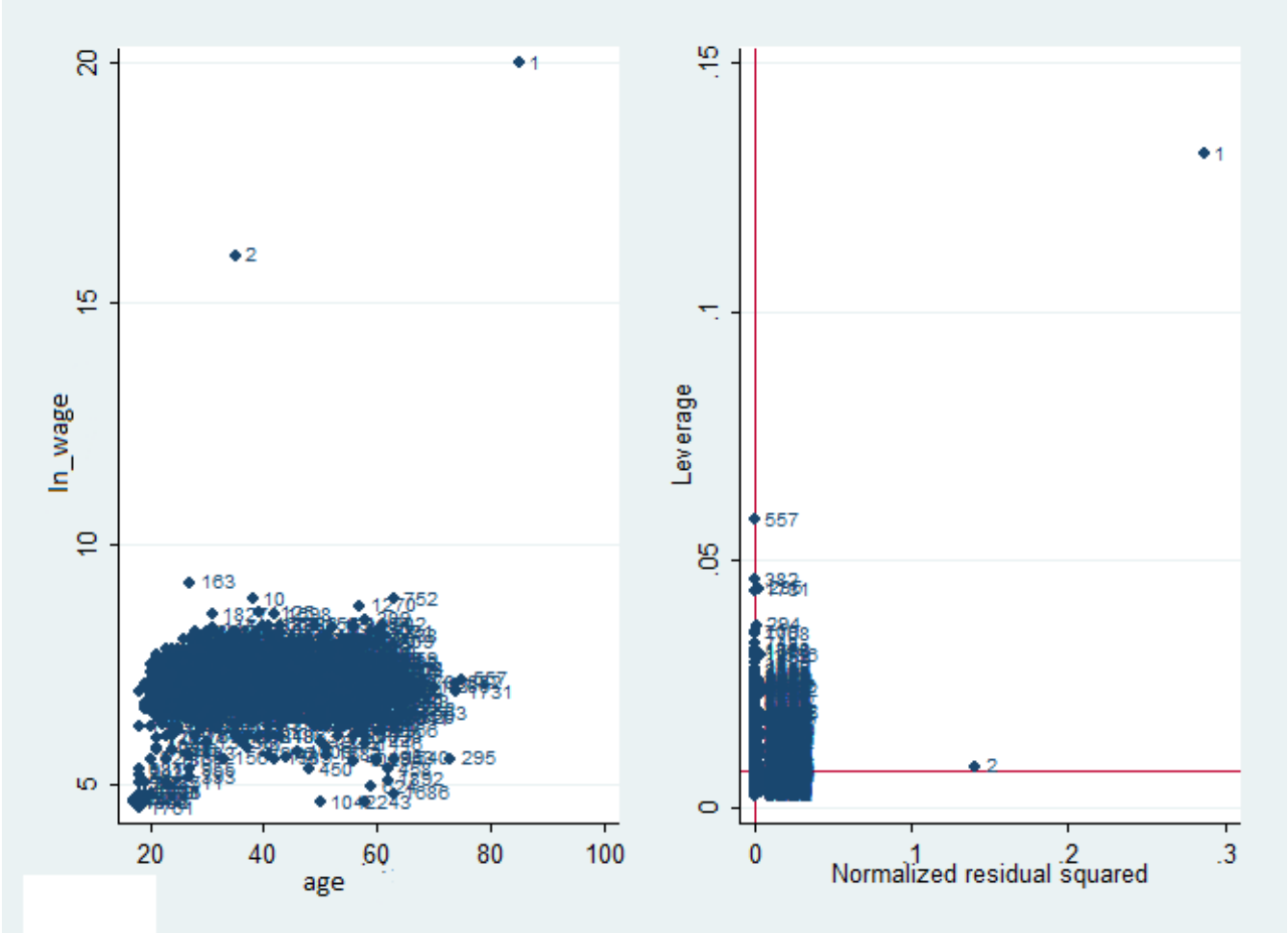
1	2	3	4	$\sum$

EXERCISE 3 Using anather data set obtained from Polish LFS researcher wants to estimate some additional models for the wage equation. Before starting the estimation procedure he calculated descriptive statistics for variable $wage$. This statistics are reported below.

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        wage |       4534    57188.11    48913.66          90      99999


count if placa==99999
2567
```

1. Explain why some observations for this variable $wage$ looks suspicious?
2. After the estimation of the regression model researcher made dot diagram for observations for ln_wage and age and plotted the graph illustrating the relationship between leverage and squared standardized residuals. Decide on the basis of these graphs which observations are suspicious and why?



3. Two additional variables were added to the regression from the previous exercise: work experience in the present workplace ($exper$) and total work experience ($exper\_tot$). Researcher estimated as well the correlation matrix for some explanatory variables included in the regression. The results of the regression and the estimated correlation matrix are reported below. What problem is present in this regression?

```
      Source |       SS       df       MS              Number of obs =    1967
-------------+------------------------------           F( 14,  1952) =  105.67
       Model |  216.967865    14  15.4977047           Prob > F      =  0.0000
    Residual |  286.280455  1952  .146660069           R-squared     =  0.4311
-------------+------------------------------           Adj R-squared =  0.4271
       Total |   503.24832  1966  .255975748           Root MSE      =  .38296

------------------------------------------------------------------------------
    ln_wage  |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      educ_1 |   .1889953   .0302267     6.25   0.000     .1297153    .2482753
      educ_2 |   .2786084   .0360692     7.72   0.000     .2078703    .3493465
      educ_3 |   .4458425   .0375599    11.87   0.000     .3721808    .5195043
       rural |  -.0512427   .0179803    -2.85   0.004    -.0865054     -.01598
         age |   .0358145   .0056632     6.32   0.000     .0247081     .046921
       age_2 |  -.0004565   .0000674    -6.77   0.000    -.0005887   -.0003243
         sex |  -.1157975   .0501914    -2.31   0.021    -.2142319   -.0173632
      w_time |   .7589422   .0441721    17.18   0.000     .6723127    .8455717
  sexXw_time |  -.1165905   .0530097    -2.20   0.028     -.220552    -.012629
    status_1 |   .0928414   .0238605     3.89   0.000     .0460467    .1396362
    status_2 |  -.1109334   .0602626    -1.84   0.066    -.2291193    .0072524
    status_3 |  -.0083049   .0442809    -0.19   0.851    -.0951477    .0785379
       exper |   .0076781   .0014517     5.29   0.000     .0048311    .0105252
    exper_tot |   .0026003    .001987     1.31   0.191    -.0012967    .0064972
       _cons |   5.566119   .1095404    50.81   0.000     5.351291    5.780947
------------------------------------------------------------------------------

Correlation matrix

             |      age   exper exper_tot
-------------+------------------------------
         age |   1.0000
       exper |   0.4875   1.0000
   exper_tot |   0.9201   0.5743   1.0000
```

4. Researcher made additional regression of the logarithm of wage on numbers of years of education ($educ\_yrs$) and obtained the following results

```
      Source |       SS       df       MS              Number of obs =    1968
-------------+------------------------------           F(  1,  1966) =   36.69
       Model |  13.7328051     1  13.7328051           Prob > F      =  0.0000
    Residual |  735.877367  1966  .374301814           R-squared     =  0.0183
-------------+------------------------------           Adj R-squared =  0.0178
       Total |  749.610172  1967  .381093122           Root MSE      =   .6118

------------------------------------------------------------------------------
     ln_wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    educ_yrs |   .1009973   .0166741     6.06   0.000     .0682967     .133698
       _cons |   6.948158   .0260974   266.24   0.000     6.896976    6.999339
```

What is the probable direction of bias of the coefficient for $educ\_yrs$ related to omission from the model of the variable $rural$?

EXRECISE..............

NAME...

EXRECISE..............

NAME...