



Übung zur Empirischen Wirtschaftsforschung

Übungsblatt 4

Please examine below OLS estimation results for the log earnings of Egyptian wage workers and answer the below questions:

Y Net basic income per 3 months in EGP

XYR Years of experience in the labor market

HRS Average number of work hours per day

ILLITERATE 1 if cannot read or write, 0 otherwise

READ&WRITE 1 if can read and write but without any certificate, 0 otherwise

PRIMARY 1 if has primary certificate, 0 otherwise

PREPARATORY 1 if has preparatory certificate, 0 otherwise

VOCATIONALSECONDARY 1 if has vocational secondary certificate, 0 otherwise

GENERALSECONDARY 1 if has general secondary certificate, 0 otherwise

DIPLOMA 1 if has diploma, 0 otherwise

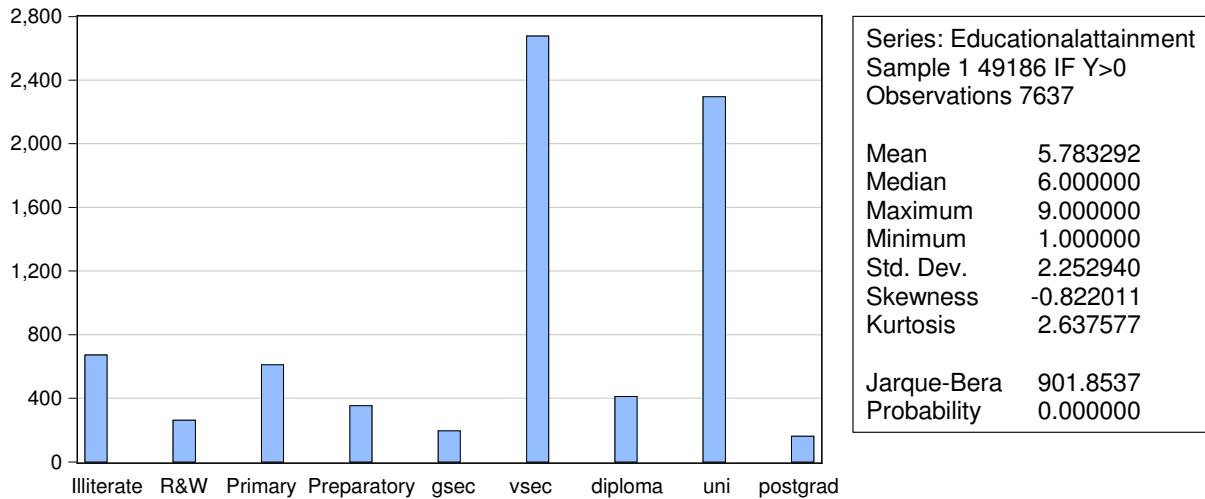
UNIVERSITY 1 if has university certificate, 0 otherwise

PRIVATE 1 if respondent works in private sector, 0 if works in the government

URBAN 1 if respondent living in Urban area, 0 if lives in Rural area

Quelle: ELMPS 2012.

Exercise 1



For which persons is the descriptive statistics for education attainment shown?
Please analyze the figure.

Given figure 1, which educational level should be used as the reference group in the earnings function estimations? Why?

Exercise 2

=====
Estimation 1

Dependent Variable: LOG(Y) Method: Least Squares
Sample: 1 49186 IF F=1 AND URBAN=0 Included observations: 518

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.995049	0.195612	35.75973	0.0000
LOG(HRS)	0.081144	0.095847	0.846602	0.3976
XYR	0.046066	0.009970	4.620637	0.0000
XYR^2	-0.000550	0.000295	-1.868122	0.0623
ILLITERATE	-0.254740	0.144515	-1.762726	0.0786
READWRITE	-0.223395	0.275788	-0.810024	0.4183
PRIMARY	0.329478	0.187283	1.759249	0.0791
PREPARATORY	-0.110726	0.206943	-0.535056	0.5928
GENERALSECONDARY	-0.149167	0.162026	-0.920634	0.3577
DIPLOMA	0.241542	0.123463	1.956388	0.0510
UNI	0.118161	0.063737	1.853883	0.0643
PRIVATE	-0.444199	0.086973	-5.107328	0.0000

=====
R-squared 0.284621 Mean dependent var 7.572090
Adjusted R-squared 0.269070 S.D. dependent var 0.703945
S.E. of regression 0.601834 Akaike info criteri1.845224
Sum squared resid 183.2754 Schwarz criterion 1.943679
Log likelihood -465.9130 F-statistic 18.30160
Durbin-Watson stat 2.042764 Prob(F-statistic) 0.000000
=====

=====
Estimation 2

Dependent Variable: LOG(Y) Method: Least Squares
Sample: 1 49186 IF F=1 AND URBAN=1 Included observations: 1078

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.602350	0.175527	37.61435	0.0000
LOG(HRS)	0.254663	0.084911	2.999154	0.0028
XYR	0.044186	0.006241	7.079753	0.0000
XYR^2	-0.000444	0.000167	-2.654622	0.0081
ILLITERATE	-0.524068	0.105685	-4.958775	0.0000
READWRITE	-0.157115	0.192645	-0.815568	0.4149
PRIMARY	0.016969	0.162597	0.104360	0.9169
PREPARATORY	-0.144592	0.141790	-1.019758	0.3081
GENERALSECONDARY	0.167528	0.129116	1.297500	0.1947
DIPLOMA	0.047306	0.077911	0.607173	0.5439
UNI	0.412538	0.044317	9.308797	0.0000
PRIVATE	0.051445	0.053611	0.959605	0.3375

=====
R-squared 0.269906 Mean dependent var 7.830539
Adjusted R-squared 0.262373 S.D. dependent var 0.693979
S.E. of regression 0.596026 Akaike info criteri1.814003
Sum squared resid 378.6927 Schwarz criterion 1.869471
Log likelihood -965.7475 F-statistic 35.82608
Durbin-Watson stat 1.862573 Prob(F-statistic) 0.000000
=====

What is the difference between the first and second estimation?

Please comment on the number of observations for both estimations. What does the difference show?

What is the average income level for each group?

Analyze the statistical and economic significance of the coefficients and the estimation quality for both models.

Compare the influence of coefficients between both models. What are possible economic reasons behind such differences?

Exercise 3

=====
Estimation 1

Dependent Variable: LOG(Y) Method: Least Squares
Sample: 1 49186 IF F=0 AND URBAN=0 Included observations: 2498

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.252730	0.099211	73.10403	0.0000
LOG(HRS)	0.131421	0.042827	3.068651	0.0022
XYR	0.024157	0.003943	6.127273	0.0000
XYR^2	-0.000289	8.83E-05	-3.273665	0.0011
ILLITERATE	-0.315025	0.042382	-7.432936	0.0000
READWRITE	-0.314101	0.058477	-5.371377	0.0000
PRIMARY	-0.277236	0.042249	-6.561926	0.0000
PREPARATORY	-0.081956	0.059352	-1.380859	0.1674
GENERALSECONDARY	0.051229	0.085555	0.598783	0.5494
DIPLOMA	0.085462	0.066071	1.293480	0.1960
UNI	0.210165	0.034438	6.102696	0.0000
PRIVATE	0.105135	0.029238	3.595876	0.0003

=====
R-squared 0.076519 Mean dependent var 7.833199
Adjusted R-squared 0.072433 S.D. dependent var 0.634945
S.E. of regression 0.611517 Akaike info criteri1.859045
Sum squared resid 929.6484 Schwarz criterion 1.887019
Log likelihood -2309.948 F-statistic 18.72633
Durbin-Watson stat 1.549272 Prob(F-statistic) 0.000000
=====

=====
Estimation 2

Dependent Variable: LOG(Y) Method: Least Squares
Sample: 1 49186 IF F=0 AND URBAN=1 Included observations: 2617

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.055708	0.108060	65.29440	0.0000
LOG(HRS)	0.232518	0.047452	4.900040	0.0000
XYR	0.029446	0.003934	7.484143	0.0000
XYR^2	-0.000317	9.22E-05	-3.442777	0.0006
ILLITERATE	-0.364090	0.050056	-7.273592	0.0000
READWRITE	-0.152275	0.075504	-2.016770	0.0438
PRIMARY	-0.207702	0.047269	-4.394072	0.0000
PREPARATORY	-0.121715	0.055468	-2.194341	0.0283
GENERALSECONDARY	0.005885	0.073629	0.079924	0.9363
DIPLOMA	0.139299	0.056012	2.486965	0.0129
UNI	0.416920	0.031419	13.26987	0.0000
PRIVATE	0.057203	0.029440	1.943050	0.0521

=====
R-squared 0.148009 Mean dependent var 8.025592
Adjusted R-squared 0.144412 S.D. dependent var 0.682074
S.E. of regression 0.630905 Akaike info criteri1.921251
Sum squared resid 1036.897 Schwarz criterion 1.948167
Log likelihood -2501.958 F-statistic 41.14049
Durbin-Watson stat 1.517792 Prob(F-statistic) 0.000000
=====

Comment on the difference between both models in terms of groups and number of observations.

Analyze the economic and statistical significance of the coefficients, in addition to the estimation quality of both models.

Compare the influence of the coefficients for both models. How is this comparison different to the comparison made for Exercise 2, and what does this difference mean economically?

How can the estimation quality of the models in Exercise 3 be improved?